View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Scholarship@Western

Western Graduate&PostdoctoralStudies

Western University

Scholarship@Western

Electronic Thesis and Dissertation Repository

4-12-2013 12:00 AM

Language in genetics research informed consent: The language gap and unrecognized miscommunication

Justin Morgenstern
The University of Western Ontario

Supervisor
Dr. Jeffery Nisker
The University of Western Ontario

Graduate Program in Health and Rehabilitation Sciences
A thesis submitted in partial fulfillment of the requirements for the degree in Master of Science
© Justin Morgenstern 2013

Follow this and additional works at: https://ir.lib.uwo.ca/etd

 Part of the Bioethics and Medical Ethics Commons

Recommended Citation

Morgenstern, Justin, "Language in genetics research informed consent: The language gap and unrecognized miscommunication" (2013). *Electronic Thesis and Dissertation Repository*. 1198.
https://ir.lib.uwo.ca/etd/1198

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlswadmin@uwo.ca.

LANGUAGE IN GENETICS RESEARCH INFORMED CONSENT:

THE LANGUAGE GAP AND UNRECOGNIZED MISCOMMUNICATION

(Thesis format: Monograph)

By

JUSTIN MORGENSTERN

GRADUATE PROGRAM IN HEALTH AND REHABILITATION SCIENCES

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

THE SCHOOL OF GRADUATE AND POSTDOCTORAL STUDIES
THE UNIVERSITY OF WESTERN ONTARIO
LONDON, ONTARIO, CANADA

© JUSTIN MORGENSTERN 2013

ABSTRACT

Informed choice is fundamentally a process of communication, reliant entirely on the tools of language. However, the meanings and understandings of words change with time, setting, and context, threatening the basis of consent. We conducted a qualitative content analysis of Canadian genetics research documents, exploring the impacts of language on informed consent. Numerous language usages were noted as potential barriers to informed consent, including language that was vague, variable, and unusually defined. Unique combinations of words were observed to generate novel concepts without clear meanings and definitions were absent or unclear. However, the ambiguity of the language was concealed by words that were simple and familiar. We conclude that a gap in communication may exist when discussing genetics, health, and disease, in that the same words, when used by different individuals, can have different meanings, and thus individuals may not fully understand each other despite using the same words.

Keywords: informed consent, informed consent documents, communication, communication barriers, language, genetics, research ethics.

ACKNOWLEDGEMENTS

It gives me great pleasure to acknowledge the tremendous support that I have received from my thesis supervisor Dr. Jeffery Nisker, who inspired me undertake a graduate degree and pursue this project, introduced me to new perspectives and ways of thinking, and continuously motivated me with his endless enthusiasm. I considered it an honour to work with Dr. Fiona Miller, who helped me find my way in the world of qualitative research and generously donated her time to assist me in the revision of my methods and methodology. I am indebted to Dr. Robert Hegele, who acted as the key informant to this study, providing insight into genetics research, acting as the first contact with other genetics scientists, and enthusiastically supporting a qualitative approach to research that is generally unfamiliar and rare in medical science. I am further indebted to Dr. Marilyn Evans, who acted as a member of my supervisory committee, reviewed my thesis, and provided important feedback. I am forever indebted to the scientists who generously volunteered the documents that allowed this research to occur. I am beholden to my many colleagues who assisted in the coding of data, development of themes, refining of ideas, and background research for this project, namely: Sylvie Galindo, Sara Dungavell, Jennifer Ryder, Sona Ghosh, Roxanne Mykitiuk, and Katharine Timmins. Without the support of all of these individuals, the writing of this thesis would have been impossible. I must also recognize the generous support of Genome Canada and the Ontario Genomics Institute for funding this project, as well as the Schulich School of Medicine and Dentistry's Summer Research Training Program for providing funding for me to pursue this research during medical school.

TABLE OF CONTENTS

LIST OF FIGURES

BACKGROUND

Medicine is increasingly being dominated by a genetic framework, which gives rise to new ways of thinking and talking about the concepts of health, disease, and normalcy (Mykitiuk and Nisker 2010). The meanings and understandings of these new genetics-impacted concepts are often markedly different from those that are still commonly employed outside of genetics (Lippman 1991, Bubela and Caulfield 2004, Hodgson, Hughes et al. 2005, Mykitiuk and Nisker 2010). Furthermore, genetic concepts are often complex, poorly defined, and difficult to comprehend (Cox, Burger et al. 2002). Thus, a significant gap in the meanings and understandings of language may exist between people working in genetics and the individuals with whom they interact, including patients and research participants (Burgess, Laberge et al. 1998). Such a gap would threaten the informed choice process, both in clinical practice and research.

The purpose of this study was to explore the meanings and understandings of terms used in the informed choice process of genetic research and recognize any concepts that may stand as a barrier between researcher and research participant and vice versa. This research is informed by three general subjects: the social impacts of genetics; informed consent and the use of consent documents; and the influences of language on communication.

Genetics

Since the advent of recombinant genetic technologies, allowing us to ‘read’ and potentially alter our DNA, genetics has taken a central seat in modern medicine (Keller

2000, Manis 2007). DNA has been described as the "book of life" (Kay 1995), the "Holy Grail" (Nathanson and Weber 2001), and the "code of codes" (Kevles and Hood 1992). Billions of dollars were spent on The Human Genome Project (The Human Genome Project 2008). It has been argued that genetics is the ideology of our time (Rothman 1998) and the gene a cultural icon (Nelkin and Lindee 2004). However, the rapid emergence of genetics has generated many complex issues (Kevles and Hood 1992, Richards 1993, Conrad and Gabe 1999, Nelkin and Lindee 2004), and the "media hype" (Bubela and Caulfield 2004), has caused misunderstandings and confusion.

Some worry that the public's adoption of genetic thinking is based on a genetic discourse that is oversimplified and extended beyond its scientific basis (Lippman 1992, Conrad 1999), and that a genetic framework is being applied to an increasing number of social problems (Mykitiuk and Nisker 2010). The media has reported on the discovery of genes for sexual orientation (Connor 1995, Pinker 2005), alcoholism (Alleyne 2010), obesity (Devlin 2008), and thrill seeking behaviour (Gower 2000), among many others (Conrad and Weinberg 1996, Conrad 1997). Lippman argues that this increasing tendency to reduce all difference to genetics, which she calls "geneticization" (Lippman 1991), results in a dangerous "genetic fatalism" (Lippman 1992). This genetic fatalism is evident in the comments made during a series of focus groups conducted by Crosseley (2002), in which she found that many individuals discounted personal responsibility for health, such as the benefits of smoking cessation, because, ultimately, is was just a matter of luck; there is no control in the face of genetic predisposition.

Although discussions of health, illness, and disability should be informed by the interaction of a large number of factors, such as social and physical environments, economics, personal behaviours, and available health services, genetics is increasingly eclipsing other considerations (World Health Organization 1948, Lippman 1991, Wilkinson and Marmot 2003, Mykitiuk and Nisker 2010). The privileging of genetics over these other factors and the complex ways in which they interact, focuses medicine and medical research on genetic "faults" (Wilson 2001), rather than the social factors that influence health.

The increasing focus on genetics is a major concern of disability scholars, who see disability as a socially constructed concept (Parens and Asch 1999, Shakespeare 1999, Wasserman, Bickenbach et al. 2005). Being in a wheelchair is not a disability when a society builds adequate ramps and automatic doors (Shakespeare 1999). Being deaf is not a disability in a community where everyone knows sign language (Groce 1985). Parens and Asch (1999) believe that basing decisions about the worth of individuals on genetic differences represents "a preoccupation with what is trivial and an ignorance of what is profound."

Another common concern is the potential for the rise of a modern eugenics (Duster 1990, Kerr, Cunningham-Burley et al. 1998, Shakespeare 1999, Nelkin and Lindee 2004). Although it is generally recognized that eugenics would not be accepted as policy, many worry about a "backdoor to eugenics" (Duster 1990). Reproductive choices are currently viewed as private and parental; however, social pressures to homogenize decisions could

unintentionally result in eugenic practices (Nelkin and Lindee 2004). Furthermore, the desire to keep reproductive choices private has limited the sense of social responsibility (Duster 1990, Parens and Asch 1999). Thus, although genetic counseling for sexual selection was once unheard of, a majority of genetic counselors indicate they would provide this service (Paul 1992).

Unfortunately, most serious discussions about the social impacts of genetics remain in academic venues, while media representations of genetics are optimistic and at times sensational (Petersen 2001, White 2001, Bubela and Caulfield 2004). Media coverage focuses on successes, while ignoring important setbacks and retractions (Stockdale 1999). Thus, the average individual approached for consent in genetics research may be aware of an overwhelming scientific optimism, but not the corresponding social concerns.

Genetics is also fundamentally changing the way we think about concepts such as health and disease (Lippman 1991, Bayertz 1998, ten Have 2003, Mykitiuk and Nisker 2010). Until recently, a person was sick when he or she looked and felt sick. Our definitions of disease were based on physically manifested symptoms. Increasingly, however, disease is becoming associated with genes. A test is run, a gene is found, and you are diseased. There are now "presymptomatic patients", who have genes that ensure one day they will become ill, but are currently healthy by traditional standards (Bayertz 1998). In most cases, however, genetics is a framework of probabilities. Thus, a category of "pre-patients or potential patients" is created for people who may never become sick (ten Have 2003). These individuals enter the medical system and are labeled as 'diseased', with all

the accompanying implications (Fleischman 1999, Hodgson, Hughes et al. 2005). They have a gene, but may never have symptoms (Bayertz 1998).

Genetics is also creating entirely new categories of disease, in which classical symptoms never develop (Issa 2002). For example, pharmacogenetics – the field of study that relates genetics to the efficacy of drug therapies – has the potential to create a new, subclinical classification of disease, in which individuals with a genetically decreased response to a drug are seen as ill (Issa 2002). Thus, adoption of the genetic framework necessitates a reassessment of the traditional boundaries of health and disease.

Consequently, genetics complicates the process of informed consent. The general hype surrounding genetics alters perceptions of the potential risks and benefits of research. The genetic framework is complex, non-intuitive, and poorly understood (Bhutta 2004). The long term genetic risks are incompletely appreciated, and are therefore difficult to adequately describe. Finally, genetics presents the possibility of unique harms, as genetic disease is not something that someone has, but rather that someone is (Shakespeare 1999). Genetics may represent a great modern medical achievement, but it is important to recognize and study the impacts of the genetic framework on medicine, research, society, and the informed choice process for research participation and increasingly for clinical diagnosis and treatment.

Informed choice

The goal of informed choice is to ensure that the decisions made by research participants are intentional, knowledgeable, and free of coercion (Faden and Beauchamp 1986, Burgess, Laberge et al. 1998, Baylis 2004, Tri-Council Policy Statement 2009). To accomplish this, participants must understand the nature of the research, its risks, and its potential benefits (Beauchamp and Childress 2001). Then, they must be allowed to form their own decisions, free of external influences (Beauchamp and Childress 2001). Neither of these steps is without difficulty, especially in the context of genetics research. The culmination of the informed choice process is the signing of the informed "consent form" both for research and clinical activities.

New genetics concepts add another layer of complexity to the already poor understanding of medical and scientific concepts (Waggoner and Mayo 1995, Bjorn, Rossel et al. 1999, Corrigan 2003, Miller, Begbie et al. 2006). Genetics relies on the difficult to understand concepts of prediction and probability, and is often concerned with future rather than present disease (Burgess, Laberge et al. 1998, Gigerenzer, Gaissmaier et al. 2008). Furthermore, genetics has introduced new classes of illness, and new forms of research (Issa 2002, Wertz 2003, Knoppers 2005, Secko, Preto et al. 2009).

Adequately describing the risks and benefits of genetic research is problematic (Hoedemaekers, Gordijn et al. 2006). Despite resounding optimism, benefits are often theoretical or abstract (Burgess, Laberge et al. 1998, King 1999, Burgess 2001). Risks are also largely theoretical, but harder to predict, as concerns often focus on social or

psychological risks (Burgess, Laberge et al. 1998, Burgess 2001, Saraiva, Anionwu et al. 2001, Wertz 2003). Many risks are unique to genetics, such as the risk of discovering new but undesired information about paternity, social history, or ethnic differences (Burgess, Laberge et al. 1998, Saraiva, Anionwu et al. 2001, Wertz 2002).

The hereditary nature of genetics complicates the concept of independent, autonomous decision making (Burgess, Laberge et al. 1998, Annas 2001, Wertz 2002, Knoppers and Chadwick 2005). Genetic information has familial impacts and decisions will be influenced by familial dynamics and perceived responsibilities (Burgess, Laberge et al. 1998, Hallowell 1999, Annas 2001, Knoppers and Chadwick 2005). Furthermore, genetic information can impact entire communities, potentially raising questions about heritage, developing stigma, or provoking discrimination (Burgess, Laberge et al. 1998, Saraiva, Anionwu et al. 2001, Wertz 2002). Research involving only a few could classify entire populations as at risk (Annas 2001).

Though informed choice is a necessary safeguard of free choice and ethical research, it may not be sufficient (O'Neill 2003, Knoppers and Chadwick 2005, Burgess 2007). Informed choice may assign responsibility to the individual, but ignore the responsibilities of institutions and society in moral practice (Hoeyer and Lynoe 2006). Informed choice alone may not be adequate to address the familial and community impacts of genetics, and genetic research frequently raises new issues such as the long term storage of specimens, multiple or unknown goals of research, and recruitment within

families, that are not adequately addressed by informed consent alone (Wertz 2003, Hull, Glanz et al. 2004, Knoppers 2005, Burgess 2007).

Informed consent documents

Consent documents are fundamental to informed consent, not just as lasting evidence of the process, but also as sources of important information for participants (Kent 1996). The information provided during the oral part of informed choice may be similar to that in documents (Sankar 2004) but can be quite variable (Bjorn, Rossel et al. 1999). Thus, the consent form acts as a consistent source to ensure adequate information is provided to all research participants. The consent form is also a lasting source of information, which is likely to be re-read (Ghulam 2006). As such, the language and structure of consent forms are important to ensure clarity and understanding.

Quality of consent forms

Some studies have reviewed consent documents with the goal of assessing the information contained and determining what important information is missing (White, Jones et al. 1996, Nolan 1999). Nolan performed a content analysis of consent documents from pharmacological research. She illustrated a lack of information provided to women about potential harms to a fetus and the procedures employed to ensure women's safety (Nolan 1999). No similar studies were found for genetic research documents.

Other studies have focused on the complexity of language and the readability of consent documents, with the idea that complex and technical language stands as a barrier to

understanding and therefore to informed consent. Considerations of average reading abilities have led to various recommendations that consent documents be written to a maximal grade level (Hochhauser 1997, Baevsky 2008, Jefford and Moore 2008). In a study of 12 consent forms from pharmaceutical studies, Hochhauser (2000) concluded that consent forms were difficult to read, with an average grade level greater than 13, too many uncommon, lengthy words, and a complex sentence structure. Hochhauser also argued that various formatting issues, such as decreasing the number of words per line and the number of lines per page, would increase the readability and accessibility of the consent forms.

Pothier (2005) reviewed the consent forms submitted to his institutional review board, and found that the average Flesch readability levels equated to the "difficult college level". Others have indicated that readability scores remain low even after institutional review board alterations (Murgatroyd and Cooper 1991, Hammerschmidt and Keane 1992). The conclusion that consent forms are too complex and difficult to read has been supported by numerous other studies (Tarnoawski, Allen et al. 1990, Lynoe, Sandlund et al. 1991, Hopper, TenHave et al. 1995, Goldstein, Frasier et al. 1996).

The issue of consent forms being complex and difficult to read has been present in the literature since at least the 1980s (Grunder 1980, Baker and Taub 1983, LoVerde, Prochazka et al. 1989). Despite convincing research in the area, readability has not improved. Over the same period, there is evidence that the length of consent forms has increased dramatically, with some now averaging over 12 pages (LoVerde, Prochazka et

al. 1989, Beardsley, Jefford et al. 2007). Within genetics, these issues are likely to be compounded, as novel concepts, new categories of risk, and complex study designs are tentatively explained.

However, the use of statistical readability formulas as an approach to rating consent documents has drawbacks. There is little evidence that these statistical values translate into better understanding (Hochhauser 1997). They emphasize short sentences and short words, but these do not necessarily make documents easier to read. Furthermore, statistical grade levels are not objective. Consent forms primarily contain biomedical language, so there is little correlation between 14 years of schooling in the arts and understanding of a medical consent form with a grade level of 14. For these reasons, Hochhauser (1997) has argued that the best assessment of consent forms results from asking readers about their understanding.

Comprehension of consent forms

When potential research participants have been studied regarding the content of informed consent forms, specific words and phrases have been focused upon and generally poor comprehension has been noted of the concepts of randomization (Waggoner and Mayo 1995, Bjorn, Rossel et al. 1999, Kodish, Eder et al. 2004, Jefford and Moore 2008), placebo (Waggoner and Mayo 1995, Corrigan 2003), and equipoise (Gallo, Perrone et al. 1995). Stead (2005) conducted a series of focus groups in which patients were presented with different consent forms. She found that common scientific terms, such as “randomization”, “blinding”, and “placebo” were poorly understood and alarming to

participants. One frustrated participant exclaimed, "'A randomized, doubled-blind study to compare durability…' Hello, hello – it's English I speak!" (Stead, Eadie et al. 2005) Furthermore, participants were very concerned that doctors would participate in such studies. Blinding and randomization were seen as disturbing, in that doctors were willfully ignorant of and uninvolved in treatment decisions (Stead, Eadie et al. 2005).

Bjorn's, noted that even with an improved consent form and a clear explanation, only 42% of participants understood the concept of randomization. They conclude that the concept of randomization is so different from the standard doctor-patient relationship that comprehension is impeded (Bjorn, Rossel et al. 1999). Similarly, other alien concepts, such as the probabilistic and predictive model of health and disease in genetics, may preclude understanding and therefore properly informed consent (Burgess 2007).

Genetics related concepts such as inheritance, susceptibility, and relative risk are largely unknown to the public (Erby, Roter et al. 2008). Although these words and concepts are familiar to and easily understood by scientists, they can be misunderstood and troubling to the research participants.

Some studies have demonstrated increased comprehension with changes to consent forms. In a randomized control trial comparing two versions of a consent form, Bjorn and colleagues showed that linguistic modifications improved both self-assessed ratings of readability and general comprehension of the research participants (Bjorn, Rossel et al. 1999). However, Davis found that rewriting consent forms from a grade 16 to a grade 7

level did not improve comprehension, despite making the forms easier to read and more appealing (Davis, Holcombe et al. 1998). Similar results have been reproduced elsewhere (Coyne, Xu et al. 2003, Flory and Emanueal 2004). Thus, quantitative measures of language are likely inadequate in assessing the complex communication entailed in the informed choice process.

Consent forms and the therapeutic misconception

One issue of informed consent that has received a lot of attention is what Appelbaum has called the "therapeutic misconception" (Appelbaum, Roth et al. 1987). Research is conducted precisely because the results are unknown. When beginning a study and enrolling participants there is little or no evidence that the studied intervention will be effective, and there is always the chance that it will cause harm. For this reason, Levine calls the term "therapeutic research" self-contradictory and illogical (Levine 1988).

Nevertheless, research participants consistently list therapeutic benefit among their reasons for participating in research. In interviews of 1,882 patients in oncology and cardiology clinics from five centers across the United States, Sugarman found that 67% of patients who had previously participated in research had done so in order to receive better medical care. Fifteen percent of all responders indicated that therapeutic benefit was their sole reason for participating (Sugarman, Kass et al. 1998). Appelbaum (2002) suggests that as many as 70% of research participants are subject to this therapeutic misconception.

The language of consent forms has been studied as a potential source of therapeutic misconception. Research participants are often inappropriately referred to as "patients" and investigators as "doctors". Rather than referring to a "treatment under study", consent documents commonly refer to "treatment" (Churchill, Collins et al. 1998, King 1999, King, Henderson et al. 2005). Within genetic research, the problem of referring to early gene transfer studies as "gene therapy" is another example of the same problem (King 1999).

Kimmelman and Levenstadt (2005) conducted a quantitative content analysis of consent documents in phase 1 human gene transfer trials. They found that the majority of forms used the term "gene therapy". They also found that 10% of the consent forms contained statements that implied that the goal of the research was therapeutic. They further noted that the format of a document may contribute to a therapeutic misconception. For example, they indicated that displaying possible outcomes of a study in list format gives the impression that all outcomes are equally possible, even though that is rarely the case.

It is important to note that this type of quantitative analysis of documents has shortcomings. It can provide valuable information about how often certain words and phrases are being used, but it is unable to identify potentially troubling words in the first place. This, of course, raises the question: how are difficult words initially identified? Some have used experience and gestalt (Waggoner and Mayo 1995). Others choose words that arise from medical or scientific vocabularies (Bjorn, Rossel et al. 1999). Asking research participants can be a productive mechanism (Stead, Eadie et al. 2005),

but it assumes that participants recognize all problematic words. However, familiar words may also represent points of miscommunication, if understood differently by researchers and potential research participants. Only qualitative analysis will be capable of discovering such gaps in language.

Language

Informed choice is fundamentally about communication. Language is used to promote understanding and request consent. Language must therefore be clear and descriptive, as any miscommunication will weaken the foundation of informed consent.

The understanding of language in my research is influenced by the work of Wittgenstein. Although earlier philosophers were concerned with finding normative rules of language, Wittgenstein recognized that the meaning of a word is in its use; a naturalistic investigation is necessary for definition. (Wittgenstein 1968) Thus, in order to explore meanings in genetic language, it is not enough to simply open a dictionary – an account of use must be developed.

Wittgenstein based his analysis of meaning in communities, defined as groups with meaningful interaction and thus a shared use of language. Each group plays a "language game" in which language becomes defined by the goals of the community. (Wittgenstein 1968) In genetics, there are many distinct communities, with varying goals, and therefore language will adopt different meanings depending on the 'winning' use of each language game.

Language is neither static nor concrete. The meanings of words change with time – for example, “cool”, “mouse”, or “spam”. New meanings are not adopted instantly and rarely eradicate previous usages (Speed 2006). Furthermore, changing language is not unidirectional. Different groups will use words differently, with the result, or perhaps the purpose, of not being fully understood by other groups (Holmes 2001).

Of course, general society is an important community of language users. Hence, for many concepts the language game is played communally, which allows general communication. However, this general capacity to communicate should not mask the fact that in many settings miscommunication is possible. The same word may be understood differently by any two people. From the standpoint of informed consent, the changing and inconsistent meanings of words could have a profound effect.

Jargon

Some language difficulties result not from differing uses, but simply from unknown words. Medical communication is made challenging by an abundance of uncommon or highly technical words that are difficult to understand from outside the field. These words are commonly referred to as “jargon”. Of course, not all jargon is avoidable. When communicating highly technical concepts, technical words may be necessary. It is generally recognized, however, that whenever jargon is used an explanation or definition should follow (Makoul 2001).

In an analysis of the language used by resident pediatricians when describing the outcomes of newborn screening, it was noted that an average of 20 unique jargon words were used per 10 minute interview. Many of these words were used more than once, resulting in a total of 72 jargon words per interview. Most residents attempted to explain some of the jargon used, but overall 83% of the words were left unexplained (Farrell, Deuster et al. 2008).

Identifying jargon words among all words that are commonly used in health professions can be a major challenge for the health professional. For example, among the list of jargon words in Farrell's study was the word "gene". As this word is encountered frequently in medical teaching, and increasingly in the media, it is difficult to recognize it as a technical concept. Farrell recognizes three classes of jargon: the highly specialized, the uncommon, and the common but confusing. He uses the example of "carrier" as a word that is commonly used, but can adopt an unfamiliar and complicated meaning within genetics (Farrell, Deuster et al. 2008). Given the probabilistic and predictive nature of genetics, it is possible that words such as "health", "disease", and "normal" have developed similarly unfamiliar and complex meanings.

The use of jargon makes medical language less comprehensible. This weakens informed consent and can potentially harm patients through mistaken expectations or increased anxiety. However, it is not just jargon that affects comprehension of medical language. Individuals who apparently speak the same language at the same level can use words differently.

Vague language

Vague language is a potential source of confusion in informed consent documents (Liss, Aspevali et al. 2004). Probability words, such as "frequent" and "rare", are interpreted inconsistently by research participants and therefore should probably be replaced by precise numbers (Sutherland, Lockwood et al. 1991, Beardsley, Jefford et al. 2007). Similarly, the designations "mild", "moderate", and "severe" are recognized as ambiguous and definitions are recommended (Beardsley, Jefford et al. 2007, Burgess 2007).

The use and meaning of the word "serious" has also been questioned. The words "serious" and "grave" were noted frequently in legislation governing access to genetic services and abortion. Consequently, Wertz (2002) conducted a survey of all members of the American Board of Medical Genetics and the American Board of Genetic Counseling, and showed a lack of agreement about how "serious" should be interpreted. Genetic conditions, such as Trisomy 21, were classified by some as "lethal", while others considered them "not serious" (Wertz and Knoppers 2002).

Within genetics, the word "mutation" is another example of vague and potentially confusing terminology. Common usage has a negative tone. Personal communication with various leading geneticists has revealed a belief that "mutation" should be reserved for situations where there is a known disease or negative consequence of genetic changes. The neutral "polymorphism" is offered as a descriptor for other forms of genetic variation. However formal definitions have commonly indicated that a mutation is any

permanent change to an organism's DNA, whether harmful, neutral, or beneficial (Condit, Archer et al. 2002, Passarge 2007). Thus, usage is not constrained by definition, allowing for confusion in communication both among researchers and with research participants.

The impact of words

Word selection is not trivial. Cognitive scientists have demonstrated that the mind is guided and constrained by the format of a symbol (Zhang and Norman 1995, Zhang 1997). Sociologists argue that behaviour towards an entity is guided by definitions or understandings of that entity (Hall 2006). Research on the medical impacts of language is somewhat limited. However, there is good evidence that the words we use affect patients and professionals alike.

Language use has been shown to impact both perception and memory development (Loftus and Palmer 1974, Zhang 1997, Zhang 2002, Heritage, Robison et al. 2007). Loftus and Palmer (1974) showed a video of a car crash and then asked the question: "about how fast were the cars going when they [smashed, collided, bumped, contacted, or hit] into each other?" The use of the word "smashed" elicited higher estimates of speed than any of the other words and also made participants more likely to falsely remember broken glass in the video.

When asking patients in a general medical practice about unmet concerns, a marked difference in response can be demonstrated by changing a single word. Patients were

much more likely to respond positively to the question “is there something else you would like to address in the visit today?” than they were to “is there anything else you would like to address in the visit today?”. The change of a single, seemingly inconsequential word drastically changed the physicians’ effectiveness (Heritage, Robison et al. 2007).

The potential impacts of language on clinical practice have been recognized (Sweeney, MacAuley et al. 1998, Guyatt, Montori et al. 2004, Stableford and Mettger 2007). Some authors argue for the use of specific terminology because of perceived benefit, or to avoid harms (Sweeney, MacAuley et al. 1998, Guyatt, Montori et al. 2004). These authors recognize the importance of terminology in keeping physicians focused on outcomes that truly matter to patients. Furthermore, many major scientific and medical associations, including the World Health Organization, the American Medical Association, and the Canadian Public Health Association, have recognized problems with health literacy and therefore emphasized the importance of plain language (Stableford and Mettger 2007).

Individual words have also been shown to be important factors in decisions to participate in research (Slevin, Mossman et al. 1995, Sugarman, Kass et al. 1998, King 1999). In a study of oncology patients’ views on research, 72% of respondents indicated that trials of a “new treatment” would be very appealing. However, when the term “experimental treatment” was substituted, only 27% responded positively (Slevin, Mossman et al. 1995). Similarly, Sugarman (1998) demonstrated that perceptions of research changed depending on how one labeled the study. For example, “medical experiments” were seen

as riskier than "medical research", but both were riskier than "medical studies" and "clinical investigations". "Medical research" was seen as providing the greatest chance of benefit. Sugarman concludes that a researcher's choice of language is an important determinant of participants' perceptions of risks and benefits, and therefore must be considered for informed consent. Language also plays an important role in potential research participants distinguishing between therapy and unknown research (King 1999). Substituting "patient" for "participant", "treatment" for "intervention under study", and "physician" for "researcher" all problematically confuse the boundary between research and standard clinical practice (King 1999).

Language and harm

In addition to the capability of language to shape perceptions and affect decisions, it is also important to recognize that inappropriate language can cause harm. (Nisker and Daar 2006) Although we all know that words cannot "break our bones", it is not true that "words will never hurt us" (Rothstein 1991). For example, words can isolate and dehumanize. As language forms a foundation of social acceptability, inappropriate words can cause social harms (Rothstein 1991). As "do no harm" (Wagley 1987) is the primary ethical foundation of health professionals, the use of potentially harmful language must be looked for and avoided.

Words can provoke unnecessary anxiety, both affecting patients' decisions and causing them undue harm (Chapple, Campion et al. 1997, Shakespeare 1999, de Crespigny 2003, Bedell, Graboys et al. 2004, Byron, Cockshott et al. 2005, Hodgson, Hughes et al. 2005).

For example, while the diagnosis “congestive heart failure” represents a treatable entity to a clinician, patients often only hear *failure,* with the suggestion of something more imminently fatal (Bedell, Graboys et al. 2004). Similarly, in interviews of patients after genetic counseling sessions, Chapple (1997) found that many commonly used genetic words, such as “syndrome”, are unfamiliar and distressing to patients.

The language used in prenatal screening has also received attention for its ability to shape decisions and potentially cause harm (Beaulieu and Lippman 1995, Benkendorf, Prince et al. 2001, de Crespigny 2003). The use of the words “baby” and “mother” in place of “fetus” and “pregnant woman” imply that it is expected that the woman will give birth and become a mother (de Crespigny 2003). This assumption, in the setting of prenatal testing, can be both harmful and unsuitably directive (de Crespigny 2003).

The disability literature provides further example of the potential harms of language, as well as the possibility of change. There is a long history of changing terms used to describe individuals with disabilities, with the most recent transitions being from “handicapped”, to “disabled”, to “person with a disability” (Devlieger 1999, Jette 2005). Using a language of “activities” and “participation” has also been suggested (Jette 2006). The emphasis of the new language is that the person comes first; that disability is only part of one’s identity (Devlieger 1999). It attempts to avoid the use of labels that privilege the trivial and devalue the individual.

In other areas of genetics, other value laden words, such as “severe”, “suffering”, and “abnormality” are commonly used (Shakespeare 1999). Byron argues that phrases such as “burden of illness” and “culprit chromosome” reinforce negative stereotypes of disability (Byron, Cockshott et al. 2005).

Hodgson (2005) conducted a survey on the use of potentially harmful language at a genetic advocacy conference in Australia. Several commonly used words, such as “mutant”, “normal”, and “disease”, as well as the practice of labeling people, were categorized as offensive (Hodgson, Hughes et al. 2005). While the respondents, being people with genetic conditions and their advocates, may be more sensitive to problems with genetic language than the general population, the same population is likely to be represented in genetic research.

The word “mutation”, although common and relatively benign in scientific dialogue, is particularly frightening in ordinary speech. In Hodgson’s study (2005), one respondent noted that, “mutant is tainted from space alien movies…” Roget’s thesaurus associates “mutation” with “freak” and “monster” (Roget's 21st Century Thesaurus 2008). It is a frightening and confusing term when applied to patients or research participants. In an analysis of genetic media coverage and a series of focus groups, Condit found that the word mutation was often presented with a negative connotation and elicited strong emotional reactions, which were not associated with the alternatives “change”, “alteration”, or “variation” (Condit, Archer et al. 2002, Condit, Dubriwny et al. 2004).

Purpose

Informed choice is fundamentally a process of communication, reliant almost entirely on the tools of language. However, language can be incomprehensible, variable, or vague, making choice less informed. Language can alter perceptions and shape actions, complicating the concept of free choice. Inappropriate language can cause harm. The purpose of this study was to explore the meanings and understandings of terms used in the informed choice process of genetics research and to recognize any concepts that may stand as a barrier between researcher and research participant, and vice versa.

The exploratory, inductive nature of the proposed research meant that we planned to analyze themes that developed from the data, rather than constraining our focus to predefined categories of inquiry. Therefore, our research questions were necessarily broad. The key questions we hoped to address included: What language is used to describe genetics research on genetics informed consent documents, and what are the possible implications of this language with regards to communication and informed consent? How are key genetics-related concepts, such as ‘health’, ‘disease’, and ‘normalcy’, portrayed through the language of genetics research documents? And could the language used in these documents impede informed consent or cause harm?

METHODS AND METHODOLOGY

General design

The goal of this research was to explore the meanings and understandings of genetics-related words in genetic research documents. In general, I aimed to be highly inclusive of any language related theme that emerged from the documents. Specifically, I was concerned with language in its communication role in the informed consent process, with a focus on its potential to result in miscommunication or harm. The general question I hoped to address was: could the language used in genetic informed consent documents have a negative impact on the informed choice process? In order to address this question, I selected a qualitative approach for several reasons. First, I hoped to recognize language uses not previously recognized, which requires an inductive approach and obviously would not allow for predetermined categories of study. Second, I felt that an emphasis on absolute numbers was inappropriate. While it may be useful to determine what percentage of consent documents make use of problematic language, such studies cannot be entertained until a better understanding exists of what language might be classified as problematic. Furthermore, certain phrases, even if infrequently used, could represent significant sources miscommunication or harm. Finally, the understanding of consent form language that I desired required an in depth reading and interaction with the material that would not be possible with the larger number of documents required for a representative quantitative study (Lincoln and Guba 1985).

Informed consent is a process of communication that includes, but is not limited to the informed consent documents. In the current study, however, I chose to limit my attention to the informed consent documents. It has been noted that what is said during an informed consent interview resembles what is present in the consent documents (Henderson, Davis et al. 2004). In fact, Sankar (2004) noted that many essential elements of informed consent that are often missed in the interview are invariably present in the consent documents. Additionally, oral information is quite variable from one participant to another (Williams and Zwitter 1994, Bjorn, Rossel et al. 1999). The consistency of consent documents facilitates their analysis and provides an important safety net in the consent process itself.

As a lasting documentation of the consent process, informed consent documents are important in their availability for review. Obviously, this is important for research, as it allows one to study informed consent without being limited to ongoing or future interactions. Furthermore, the ability of consent documents to be reread by patients and study participants increases the importance of the information and language they contain (Ghulam 2006).

An assumption might also be made about the language present within consent documents. Unlike clinical interactions, consent documents are reviewed many times, both by their authors and by research ethics boards, before they are seen by patients. This may result in more carefully selected language than is possible in verbal communication. Problems that are apparent in such a formal context may be seen as more problematic than those

observed in clinical conversations. However, it must also be recognized that the production of consent documents represents an artificial scenario, heavily influenced by legal concerns, which may create unnatural or confusing language that one would not encounter in oral communication.

Finally, documents provide important access to the context of the consent process, as original grant applications, research documents, and publications can all be included to provide a richer understanding of the research and consent process.

Data collection

The focus was the identification of funded genetic studies involving human populations and the collection of all documents used within each study: grant proposals, surveys, information letters, consent forms, as well as the ensuing publications. Studies were selected to provide maximal depth of information pertaining to the language of health, disease, and normalcy in genetic research, especially considering the importance of language to the process of informed consent. Studies of interest were those working with adult onset conditions in human populations. Studies looking at congenital and early onset genetic conditions were excluded from the present analysis due to the additional complexities of communication and consent in pediatric populations. For the same reason, studies were excluded where the patients could have a lack of capacity, such as studies of psychiatric conditions, and substitute decisions makers may be involved. Studies examining populations in the developing world were also excluded due to concerns that the context of consent and understanding of genetic health and disease

would be too far removed from our own. Finally, I excluded studies that did not involve human contacts, such as those using existent registries or data mining techniques. After identification of appropriate grants, lead authors were invited to participate by the principal investigator of the Genome Canada grant that funded this research: Dr. Robert Hegele. The involvement of Dr. Hegele, a prominent and well respected Canadian genetics researcher, as our key informant and primary contact during recruitment was felt to be essential to achieve buy in from genetics researchers for an ethics project using unfamiliar qualitative methodology and designed to analyze potentially sensitive topics.

The analysis began with a single large genetics grant that encompassed multiple projects which fit our inclusion criteria. The grant was chosen partially on the basis of convenience, as the primary investigator was known and eager to participate in the research. However, the research itself also played an important role in the decision. The grant included a diversity of genetic research. Both common and rare conditions were represented. Simple Mendelian inheritance was studied along with complex genetic interactions and newly discovered forms of genetic variation. I felt that the inclusion of such a variety of types of research was important, considering the general exploratory nature of this project. As the research progressed, further genetic grants were included to expand certain themes, until a theoretical saturation was attained. (Admittedly, this was a difficult point to define, as the emergence of new themes always led to new questions. Also, defining the study as exploratory, there were always further grants that could have been explored. Therefore, what truly defined the endpoint for data collection was the

concept of diminishing returns, combined with practical considerations of the desired scope of a master's thesis.)

Only operating grants were considered. While it would be interesting to explore the use of language in all grants, in this preliminary study we felt it was important to limit ourselves to those studies that involved larger populations and would likely be viewed as leaders in their fields. Due to the rapid development of genetic knowledge, technology, and the associated language, only the most recent studies were sought. Recipients of grants awarded in 2007 (the year that data collection began) were approached first and older studies were subsequently requested until data saturation was achieved. CIHR and Genome Canada were relied upon as sources of grants to limit the variability in scope and regulation that exists in the extensive field of genetics. From these funded grants, I collected the grant proposals, surveys, information letters, consent forms, and any ensuing publications. After being offered a copy of a REB application by one participant, such documents were subsequently added to our list of requests. All documents received were placed in an electronic database that was subsequently loaded into the qualitative data analysis software NVivo 7™.

All documents were generously contributed by highly respected genetics scientists. By using such a select group of researchers, we were assured that the best interests of patients as research participants were always the primary goal of these documents. Therefore, we knew prior to starting our analysis that any complicated language that might be identified would be unintentional. We anticipated that such language may exist,

simply because research such as this had not previously occurred. That these researchers were interested in participating in the project indicated their commitment to their research participants in general and particularly informed choice, as well as their willingness to subject their own documents to critique to prevent potential harm to their research participants.

In total, 33 scientists were sent an invitation to participate through our principal contact, and 16 responded, providing their current research documents. Two individuals were unable to provide electronic copies of their documents and another indicated that the original consent forms were no longer available, and so these grants were not included. Individual grants were purposefully selected from the remaining 13 grants based on a review of their content, and analyzed one at a time until analytical saturation was reached. The end result was a group of seven grants purposefully selected for their depth of information, presumed importance within genetics research, and relevance to the larger issues of communication and consent (Miller and Alvarado 2005). It should be noted, however, that aside from the exclusions mentioned above, grants were not selected or excluded based on the focus of their work. Thus, the choice of what diseases, conditions, or traits were important enough to be studied was left in the hands of the geneticists doing the research.

Choice of Methodology

I conducted a qualitative content analysis to explore the uses of genetic language surrounding the concepts of health, disease, and normalcy, as expressed within the

collected research documents (Hsieh 2005, Mayring 2002, Miller and Alvarado 2001, Morgan 1993) A theoretical distinction has previously been made between content analysis and context analysis. The former approaches documents as independent sources of knowledge and attempts to determine what information they contain. The later approaches documents as part of a social structure and uses them to answer questions such as how and why they were developed (Miller and Alvarado 2005). Practically, though, this distinction is difficult to maintain. A document's context cannot be fully determined without knowledge of its content and knowledge of its content is valuable only insofar as it relates to its context. My primary task was the identification of the various usages and understandings of the language of health, disease, and normalcy expressed within the documents – the content. However, this was simultaneously accompanied by an analysis that considered how and why such usages are employed and what impacts they may have – the context. Thus, the following represents the intertwined processes of content and context analysis, with the overarching goal of an understanding of the genetics language used in the studied genetics research documents.

Content analysis can further be subdivided into that which focuses on manifest or apparent data and that which uses latent or inferred data (Graneheim and Lundman 2004). Again, this analysis required the use of both. I hoped to describe how words were being used, in what combinations, and in what context. Additionally, I strove to examine how terms were understood or defined. Aside from the rare case of a dictionary or glossary, the meaning of words is wrapped up in their context and accessible only through an analysis of latent data (Wittgenstein 1968, Holmes 2001, Condit, Archer et al. 2002). In

fact, documents can provide important clues about how words are operationally defined without ever using them. (It would be possible to write a treatise on the meaning of health without ever using the word “health”.) Thus, the research goals necessitated the use of both latent and manifest content analysis.

The qualitative analysis of documents differs from analysis of other sources in one important manner: the data to be analyzed is generated prior to its being collected (Miller and Alvarado 2005). While the researcher collecting data from interviews, focus groups, or observations controls the production of data and knows the reasons for its creation, the researcher who studies documents must strive to account for the unknown motives that drove the creation of the documents. This can, and should, be accomplished both at the level of document selection and within the analysis.

As described above, documents were selected to ensure they were rich sources of information and products of an understood context. Grants submitted to funding agencies such as the CIHR and Genome Canada, as well as the documents within, are created primarily to facilitate the research process, gain funds, and ensure ethical research practices. Thus, through the selection criteria, I was able to avoid documents created for unknown or potentially dishonest reasons. The context of document creation was further considered as part of the analysis. Individual documents were coded according to type, allowing important differences in motive - as occur, for example, between consent forms, grant applications, and surveys - to later be readdressed in the process of theory testing

and model development. Additionally, theoretical memos were written to address the potential motives behind each individual document type.

Analysis

The approach to analysis and coding that was employed derived from the work of Strauss and Corbin (1998) and that of Lincoln and Guba (1985). Three analysts coded the documents using inductively developed categories to emerge and determine themes of language use. First, all documents within a case (a single research project) were read in their entirety to help the researchers immerse themselves in the data and gain a sense of the whole. Subsequently, documents were read line by line as part of the process of open coding. Descriptive codes were applied to the smallest unit of text containing the overall idea, with an emphasis on deriving the code's working title from the text (Strauss and Corbin 1998, Hsieh and Shannon 2005). As data emerged and similarities became apparent, the process of axial coding was integrated to further describe developing categories. Codes evolved using the constant comparative method, in which data to be coded was compared with all data previously coded in the same category, allowing for the modification of code descriptions and reclassification of older data as necessary (Lincoln and Guba 1985). The coding process was reliant on and supplemented by the frequent use of analytic memos. Memos are important to record first impressions, develop categories and theory, and function as an audit trail (Lincoln and Guba 1985, Catanzaro 1988).

Data that did not initially seem to fit into an existent code, nor warrant the creation of a new code, was temporarily placed in a miscellaneous category to be readdressed as coding progressed, ensuring that data was not lost (Lincoln and Guba 1985). However, the use of the miscellaneous category was intentionally limited. The existence of ambiguous data can be thought to indicate difficulties in the developing coding structure and conflicts in analysts' thoughts. Thus, the discovery of ambiguous data was followed by a break from coding to allow for the writing of memos, both about why this particular data was difficult to code, as well as about the how existing coding structure excluded the difficult case (Lincoln and Guba 1985).

The use of three analysts was intended to reveal more about the data through multiple viewpoints and act as a form of triangulation as differences between coding structures were collaboratively reviewed. The independent coding structures were reviewed primarily at the levels of open and axial coding. The differences and similarities in coding structures, including specific data coded, grouping of codes, and naming of codes, were compared. Similar categories were combined, using differences to refine category descriptions. Analytical memos were written collaboratively to explain the reasons for the differences in coding structures, to assess for bias, and to ensure that each analyst's codes were adequately grounded in the data. Divergent observations were noted, discussed at length, examined through analytical memos, and ultimately included in the coding structure if they were felt to be adequately grounded in the data. The principal investigator (Justin Morgenstern) was responsible for integrating and refining the multiple coding structures and completing the process of selective coding.

Although documents were labeled according to type to allow for future questioning of like documents, throughout the initial stages of analysis a case based approach was adhered to. Thus, all documents from a single project were examined together as part of a larger whole. This approach was necessary in order to understand the documents, fully grasp their context, and approach the analysis of meaning in an intelligible manner. Documents are fundamentally intertextual, with contained information relying on and developing from other documents (Miller and Alvarado 2005). This is especially true of documents created as part of the same project. For example, a consent form is meaningless without the attached information letter, and both are difficult for a researcher to interpret without a grant describing the proposed research. Furthermore, the depth of understanding essential to qualitative research is based in large part on the development of a rich context (Catanzaro 1988). Significant perspective would have been lost had documents been approached individually. Finally, aside from the rare document that contains a glossary, definitions are never explicit. The meaning of words is derived through their use and context (Wittgenstein 1968, Holmes 2001, Condit, Archer et al. 2002). Thus the inductive process of reconstructing understandings of words is greatly aided by increasing the number of examples and spreading the breadth of context. That being said, it was noted that each 'case' likely had multiple authors, including numerous scientists as well as individuals who specialize in the writing of documents such as consent forms. Thus, the potential for multiple intended meanings was anticipated and incorporated into the analysis.

Throughout the process of data analysis, several techniques were employed to help ensure that constructions were adequately grounded in data, credible, and fully developed. Analysts immersed themselves in the data, providing scope through prolonged engagement and depth through persistent observation (Lincoln and Guba 1985). Depth of understanding was further enhanced through purposeful sampling and the development of a rich context through the case based approach to analysis. The use of three independent analysts provided a variety of perspectives and allowed for the triangulation of theory and the generation of multiple convergent constructions (Lincoln and Guba 1985, Catanzaro 1988). Reflective memos were written whenever new ideas were developed and whenever data was encountered that did not fit within existent coding structures, to ensure all themes were fully developed and grounded in the data (Strauss and Corbin 1998, Lincoln and Guba 1985). Peer debriefing, both among analysts, and with an independent disinterested peer (an individual not involved in the coding process), was important in testing hypotheses, unearthing and exploring bias, and "keeping the inquirer honest" (Lincoln and Guba 1985). Finally, using the built in functions of NVivo 7, an audit trail was created to allow the process of analysis and resultant constructions to be reviewed (Catanzaro 1988).[1]

[1] As is explored in Theme 1 "The difficulty with presenting qualitative data ethically and anonymously", the invitation to participate in this study included a promise of anonymity for the research participants. This limits the practical value of an audit trail, as my raw data is not anonymous. Therefore, it is unclear who should be allowed access to this data, and in what context.

THEMES AND SUBTHEMES

Five major themes emerged during the coding process and are outlined below. Theme I explores the tension between the ethical requirement that research participants remain anonymous and the rich context required to adequately present qualitative research results. Theme II explores the difficulty encountered in coding and understanding the language used to describe genetic phenomena. Theme III considers the absence of explicit definitions in informed consent documents. Theme IV discusses a unique representation of "family" in the studied genetic research documents. Theme V describes and explores the impacts of a linguistic tool which I have named "conceptual meshing".

I. The difficulty with presenting qualitative data ethically and anonymously
    a. Identifying data is difficult to identify
    b. It is unclear how best to anonymize data
    c. The consequences of altering presented data
II. Complex, vague, and variable language
    a. The language of causation
        i. Direct causation language
        ii. Susceptibility language
        iii. Variable language use
        iv. Vague causation language
        v. The role of external references
    b. Descriptions of the causative agent
    c. The interchangeability of condition labels
    d. The phenotype
        i. Condition labels are empty
        ii. Normalization of the phenotype
        iii. New classes of illness
III. The lack of explicit definitions
IV. The genetic family
V. Conceptual meshing
    a. The "disease individual"
    b. Other meshed individuals
    c. "The X gene" and other meshed concepts

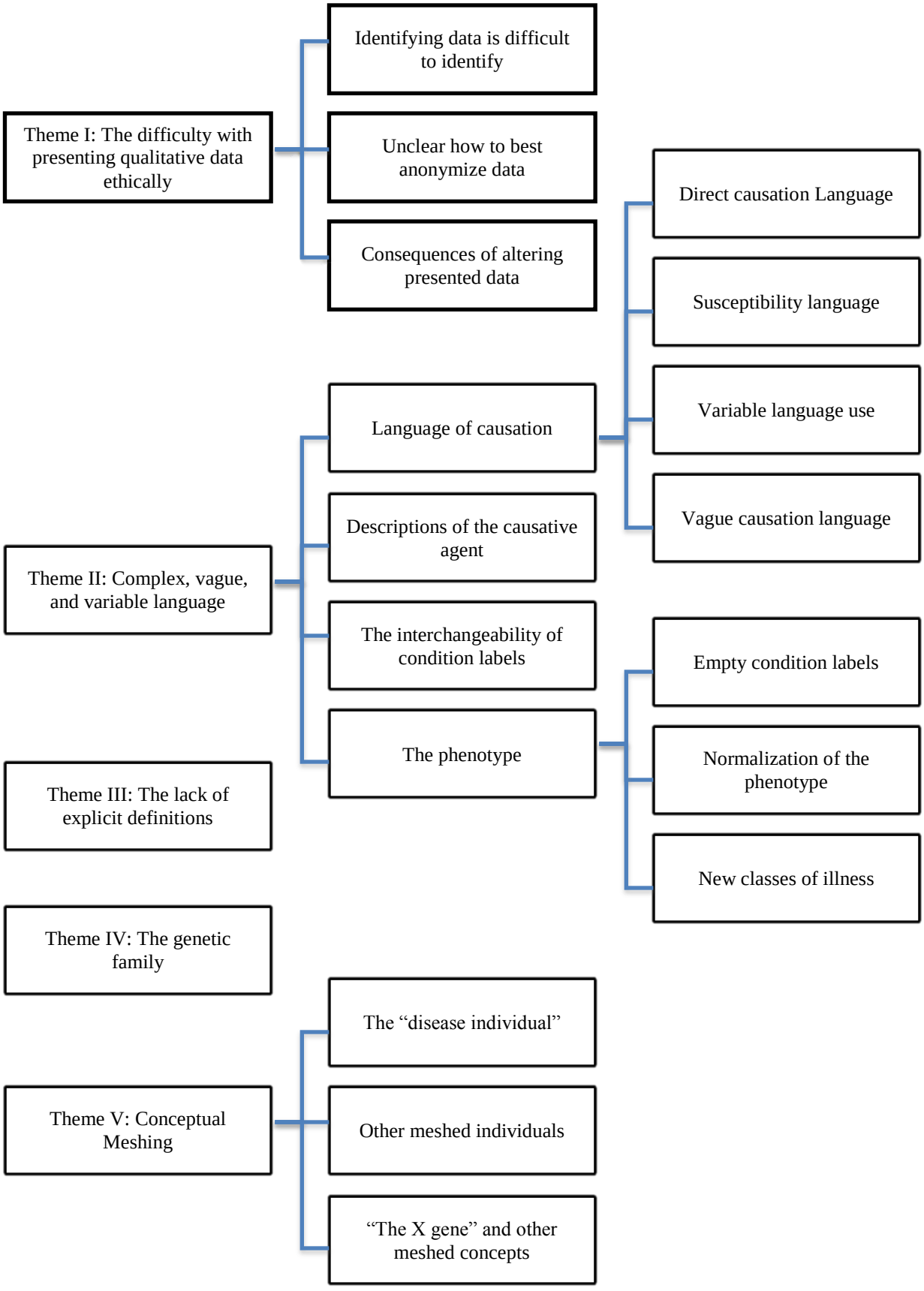

Figure 1. Major themes and subthemes

THEME I. THE DIFFICULTY WITH PRESENTING QUALITATIVE DATA ETHICALLY AND ANONYMOUSLY

In the process of preparing my results for presentation, I became aware of a tension between the necessity to keep data anonymous and the importance of direct quotations and rich context in qualitative research. To ensure the anonymity of research participants, it was necessary to replace details in some quotations with generic placeholders. For example, I used "[condition]" to replace any named condition and "[location]" to mask the name of a place. Initially, this process seemed to work. It made the data anonymous while, I believed, preserving the essence of what was being said. However, as the process of writing continued, many problems became evident: it was often difficult to determine what details required elimination; it was unclear how quotations should be altered; changing quotations influenced how the results were presented; some data could no longer be presented at all; and with each substitution important but potentially concealed assumptions were made. That my purpose was to analyze the use of language made the need to alter language especially problematic. The words were not just a tool to present data, they were the data.

Identifying identifiers

With a biomedical background, I assumed that recognizing identifying data would be simple. Names, places, telephone numbers, birthdates, etc. are all relatively easy to spot and eliminate. However, in preparing the data for presentation I quickly realized that such

superficial identifiers were a minor worry. The richness of qualitative data meant that participants were identified not just by name, but also through the details presented.

The difficulty in identifying identifiers stems both from the need to recognize data as potentially identifying and from the need to decide how much anonymization is sufficient. Some data, such as information about the location of studies, was easily recognized – possibly because the same data would be anonymized in biomedical research – and it was also easy to eliminate, as location is highly identifying but of limited value to the study.

Other potentially identifying information was often overlooked on a first reading. It is not clear to me why, but many words that could identify the research participants initially seemed innocuous. For example, in the following sentence, the word "women" was not initially recognized as potentially identifying data, despite the fact that it obviously provides information about the study population, and therefore potentially about the genomics researchers involved.

> *We will analyze* [] *from unrelated women who are at high risk of carrying a susceptibility mutation…*
>
> - Project 7, Grant Application, Line 153

I found that recognition of potential identifiers required extensive time and concentration. A brief skim through the data would not be sufficient. Indeed, even with a detailed focus

on anonymization, I cannot be sure I have identified all identifiers. Presumably, this process becomes even more difficult when using other forms of qualitative data. I imagine it is extremely difficult to know what details of an individual's life story are unique, allowing identification by family and friends.

The word "women" is also an example of identifying data that may not truly identify. Obviously, as the quotation appears here unchanged, I decided it should not count. The knowledge that women were involved in the study certainly indicates a certain subset of all genomics research. However, there are so many studies that involve women, either exclusively or as a subgroup, that such knowledge on its own is surely not identifying.

Similar considerations were required when contemplating the methodological language in these documents. Specific methods or technologies mentioned could identify the grant. However, some techniques are so commonly employed that they would be powerless as identifiers. Without the appropriate knowledge of genomics, though, these common procedures are difficult to differentiate from the more unique techniques. Thus, in this study, my paucity of genomic knowledge necessitated the general exclusion of methodology identifiers, unless I could be certain that they were commonly employed. In general, it seems that a certain degree of expertise will always be required for anonymization – one must understand the field and its participants to truly know what data could identify.

My approach to the difficulty of identifying identifiers was to consider every word as situated along two spectrums: 1) from minimally to highly identifying and 2) from unimportant to central to the research. A representation of this scheme is presented in figure 1. Each point represents a potential word. The line represents a theoretical cutoff for including a word, though, as will be explored below, there are numerous problems with such a concept. This system was not perfect, but it allowed me to consider the impacts of keeping or removing each word.

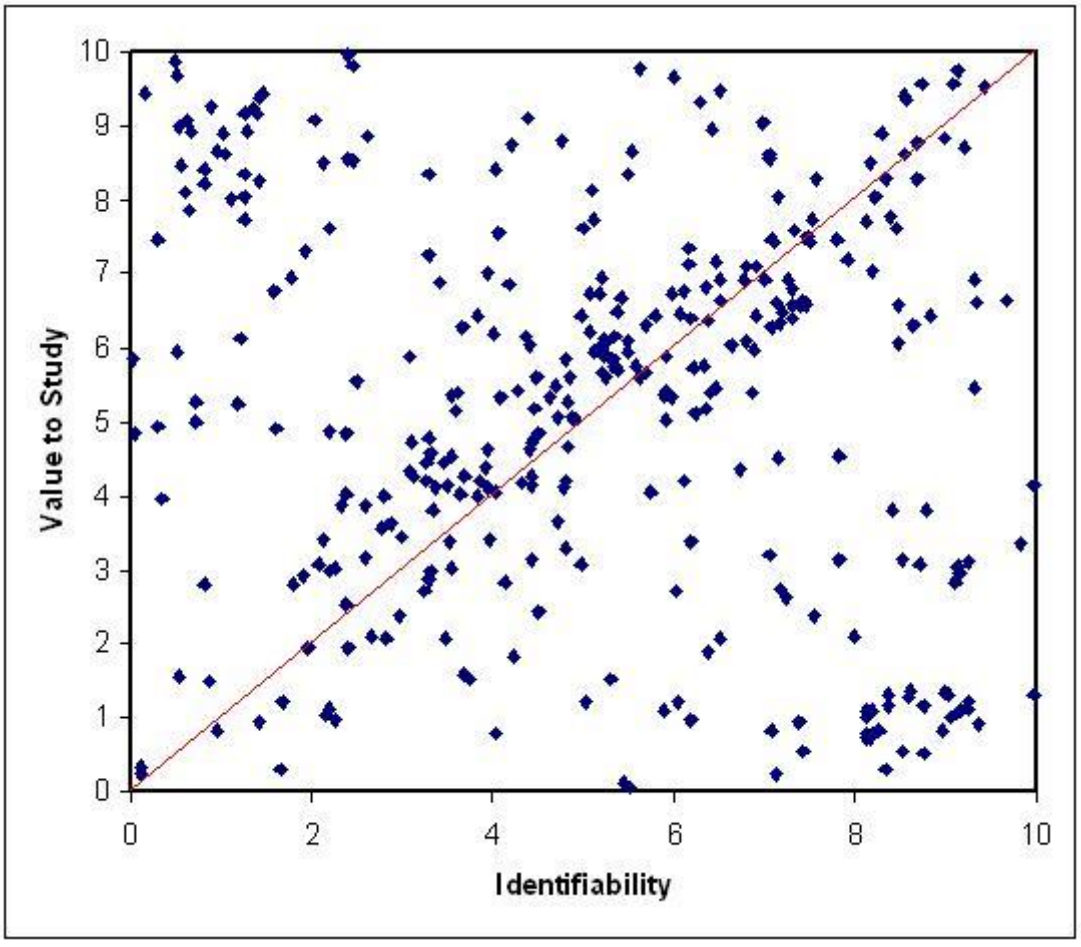

Figure 2. A representation of the properties of words considered in anonymization decisions.

The extremes of these spectrums are relatively easy to deal with. If value is high and potential to identify low, the data should be kept. If value is low and potential to identify high, the data must be eliminated or altered.

When the potential of data to identify was very low, my instinct was to favour maintaining the original language, even if the value of specific words was also considered low. Although the individual words may not have been central to a theme, they provided important contextual information and improved the flow and coherence of sentences, improving comprehensibility and maintaining the vitality of the language. Furthermore, it is often difficult to know what data will become interesting or relevant, either to future research or to the audience. Therefore, I aimed to preserve as much of the original language as seemed possible while still ensuring the anonymity of my research participants.

On the other hand, when the potential of data to identify was very high, obligations to research participants outweighed the value of the data, even when that data was integral to theme development. My initial invitation to participate included a promise of anonymity in presented data, thus my obligation was clear: despite the loss of important data, anonymity had to be the priority.

Unfortunately, most decisions were not so simple. The majority of words were situated somewhere in the middle of each spectrum. They were valuable and informative, though perhaps not central to a theme. They were also identifying, in that they provided

information about the participants, but did not directly identify. I was left with no guide in deciding how the data should be anonymized. Each word had to be considered individually, weighing its value to the research against its potential to identify.

The above treats each word as if it exists independently, and therefore over-simplifies the process of anonymization. In the presentation of qualitative data, value and identifiability are cumulative of all words. When presented independently, words may be rated as 'low value' to the study. However, whenever a word is removed, the context, flow, and coherence of the presented language are diminished. In combining to form phrases, sentences, and entire documents words become more valuable than they ever are on their own.

The same is true of words as identifiers. Independently, a word may rank low on the 'identifiability scale', but when combined with other seemingly innocuous words the cumulative effect can be highly descriptive. I argued above that the word "women" has minimal potential to identify. However, if information was also provided about the population's age, ethnicity, location, symptoms, treatment, or genetic make-up, the research population may be easily identified. Importantly, such additive information need not be in proximity. Presented data can additively identify when presented in different quotations, different chapters, or even different publications. Although, practically speaking, it may be more worrisome to see all identifiers in a single sentence, if one strives for true anonymity distant accumulation must also be considered.

Additive identifiability makes the presentation of qualitative data far more complicated. Some limit to the degree of allowable identifiability must be determined. Somehow, the researcher must decipher the impact of combining words, as each set of words will be uniquely valuable as well as uniquely identifying. The researcher is no doubt the best judge of the value of words. On the other hand, determining identifiability may be beyond the researcher's ability. As we have seen, the task often requires specific expertise. One needs to know not only how identifying each detail is on its own, but also how details may combine to identify. In my experience, the complexity can be overwhelming.

One possible solution is to present the data either to the study participants or to the studied community. If individuals intimately familiar with the local context of data collection are unable to identify participants, it is probably safe to claim anonymity. Unfortunately, if the presented data is not anonymous, research participants have been identified precisely where anonymity is most important: among individuals who know the research participants (Nespor 2000). Furthermore, such a system may verify anonymity, but cannot determine if the researcher has been overzealous, removing data that could have remained. This technique also does not account for an audience that may actively try to identify participants, potentially using external resources, such as journalists.

Additive identifiability requires that the researcher recall all previously presented data. If we are concerned with true anonymity, then even data presented in different articles or

different formats must be considered additive. However, the depth of qualitative research means that results are often presented over long periods of time, as the researcher develops and revises theories, making attention to such details difficult.

Finally, the additive effect of qualitative data means that, after data has been presented, decisions must be considered final. Presented data is 'out there' for comparison. In my decision to include the word "women" over other potentially additive identifiers, I have limited my future possibilities. Reflection may provide new insights, but if any data combines with "women" to identify, I will not be able to present it.

Presently there is very little guidance available to make such decisions. I was greatly aided by the imagery of many overlapping circles – a Venn diagram. Highly identifiable data would be a small circle – representing a small possible population. Larger circles represent a larger possible population. As many circles overlap, the potential population is drastically limited. That is, to know that a study population is male only eliminates 50% of the population, but when information about age, location, occupation, etc. is added, the possible population is quickly limited. Obviously, then, not all such details can be included. It is up to the researcher to choose the 'circles' that are most important to the research, and then anonymize the rest.

How to anonymize

After identifying data to be anonymized, another difficulty lies in deciding how it should be masked. Many options are available to anonymize individual words. One could simply

black words out, leave blanks, use codes derived from the words, use descriptors, or try to find distinct but equivalent words (Rock 2001). These methods vary in both the degree of anonymization and the retention of data. The best method will depend on the type of data and the goals of the project.

Simply removing words, or replacing them with incomprehensible codes, alters sentence structure and limits readability. However, choosing a substitute word is difficult. For example, in anonymizing interview participants, one would use “male #1”, “female”, “nurse”, “doctor”, “child”, or “teacher” as descriptive placeholders. These terms provide potentially valuable context to the audience, but also retain the ability to identify. (Consider the use of “doctor” in a study of a rural community with only a handful of doctors.) On the other hand, the use of a more general place-holder, such as “individual 1” or a blank, is certainly anonymous but loses valuable information. The choice of replacement may also impact on the structure, complexion, and comprehensibility of the quotations presented.

In this study, the choice of how to replace specific condition titles was particularly difficult. A vast array of biology is studied in genomics, and finding words to replace such variety is problematic. The first time I encountered this problem, I was lucky. I had been stuck trying to determine the best placeholder(s) for the sentence:

> *We are conducting a research study looking at genetic contributions to* [_____], [_____] *or* [_____] ...
>
> - Project 1, Consent Document 1, Line 110

The three phenomena described were quite varied and I was unsure how to proceed. However, in this case, returning to the original text provided an answer. In the next sentence, the author refers to all three phenomena as "conditions".

> *Earlier studies have suggested that some genetic differences in special populations are linked to* ***these conditions****.*
>
> - Project 1, Consent Document 1, Line 111

Consequently, I felt comfortable in choosing "[condition]" as a replacement. The sense of the sentence was still altered, potentially making the results less clear, but at least the replacement word arose from the author's own lexicon.

Unfortunately, I was routinely required to choose a place-holder without any clues from the text. I had to decide whether a certain condition was negative enough to count as a "disease", or if another word fit better. This was impossible, given that the very reason for undertaking this study was to explore the meanings of these words. Thus, I felt powerless to choose the 'correct' word, and instead focused on the potential impacts of my choice. Some words, like "disease" and "disorder", have been described as pejorative,

especially in the context of genetics, and therefore were avoided (Hodgson, Hughes et al. 2005). Other words just seemed too specific.

Ultimately, I decided to replace all specific condition names with "[condition]" because it was the most neutral and encompassing term I could find. Unfortunately, the tradeoff is that "condition" is vague (being defined as both a state of health and of disease) and uninteresting (The American Heritage Dictionary 2004). Consequently, sentences with this substitution assume a tone of bland neutrality that is not present in the originals. Furthermore, the repetition of the same substitute may provide the reader with a sense of consistency that was never seen in the original language.

When replacing specific condition titles, it was also necessary to decide whether to label each condition uniquely. By using only one word, "condition", it would be impossible to observe how language differed between conditions. To retain such information, a further code would have to be developed. For example, "condition 1" and "condition 2" could have been used to replace distinct condition names. For this study, I decided a single place-holder was the best approach. My analysis did not find any major differences in the language used to describe specific conditions, so I did not feel that multiple place-holders would be valuable. Furthermore, the number of named conditions across the seven projects was so large that organizing specific codes would have been difficult, and remembering them nearly impossible.

Sometimes terms were so technical or specific, I decided the best that I could do was replace them with "[]". This indicated, to me, that a specific or technical word was being used for which I could find no alternative. However, an assumption that underlies this project is that all words are specific and have individual meanings. Thus, the symbol "[]" might have been more fitting throughout. Ultimately, however, I decided the use of words as place-holders, where possible, was necessary to maintain the comprehensibility of the sentences presented.

In addition to the words and language used, the anonymity of presented quotations is also influenced by reference to its source. The first step that I took in the anonymization of my data was to replace the titles of grants with generic titles, such as "project 1". In choosing this placeholder, many of the same issues discussed above applied. For example, some contextual information would have been added by including the year that the grant was awarded in the label, but there are so few genomics grants in the Canadian context that such information may have been too identifying.

Quotation labels allow comparison of multiple quotations from the same project, providing context but also permitting additive identification. The words "young", "male", "lung", and "infertile" may individually be equivocal, but together point strongly towards a study of cystic fibrosis. By hiding the source of quotations, words appearing separately cannot be combined, and therefore are less powerful as identifiers.

Despite the anonymity benefit, I decided that labels of quotation source were important in this project. Although most presented quotations are single sentences, documents build meaning that extends beyond individual sentences. By presenting the source of quotations, the audience can develop a greater sense of each project, despite having never seen any document in its entirety. Source was also important in the revelation of a major theme of my work: variable language use. It was important that I was able to demonstrate that the large range of language I encountered was often contained within a single document.

Furthermore, an indication of source is an important component of the "audit trail" that allows assessment of my work (Catanzaro 1988, Strauss and Corbin 1998). The source code is the remaining link between final product and raw data, which will allow interested parties to examine my work. (However, as my raw data is not anonymous, an interesting side issue exists in determining who should be granted such access.)

The consequences of altering quotations

Changing the language of quotations frequently muddied the presentation of themes, which were, after all, about the language being presented. For the most part, I found that altered quotations were still capable of illustrating the themes. However, I was struck by a sense that the presentation was always weakened. Consider the quotation used to introduce the theme of conceptual meshing:

Some remaining subgroup of **[condition]** ***individuals*** *will have …*

- Project 7, Grant Application, Line 71

The altered language is the same language that I want to bring to the attention of the audience; it is the language from which the theme developed. However, a specific name has been replaced with the mysterious "[condition]" and the reader is left to ponder what kind of individual this might have been. Although I am able to illustrate this novel use of language, the essence of the words is lost. What makes this language so intriguing is the incredible emotion it could provoke. There is an enormous difference between a "lung cancer individual", a "diabetes individual", and a "hemorrhoids individual". The general concept is preserved, but the imagery and feeling of the sentence is lost.

In addition to such emotional content, the removal of specific condition titles also removed important information. For example, specific genetics titles often contain words such as "familial" or "hereditary". When such a title is replaced with "[condition]" integral information about the condition's presumed familial nature is lost. A "familial condition individual" is an intriguing, complex concept that is forever lost behind the veil of "[condition]".

On the other hand, when the altered language is peripheral to the theme, it is tempting to claim that changes are undisruptive. Unfortunately, I found such changes to be equally troubling. For example, consider these quotations from the discussion on causation language:

Below you will find a brief explanation of our study to identify the cause of [condition].

- Project 2, Consent Document 1, Line 5

These studies will include analysis for genes that affect the risk of developing [condition].

- Project 7, Consent Document 3, Line 30

Title of Project: Characterization of genes involved in [organ] *development and* [condition]

- Project 3, Consent Document 2, Line 6

As before, the specific condition titles in each sentence have been replaced, but now the words of interest are “cause”, “risk”, and “involved”. Initially, these peripheral changes seemed to leave essence of each statement intact. However, closer examination reveals that the problems associated with “[condition] individual” are still present. Again, the sense or tone of a sentence about the “risk of diabetes” is drastically different from one about the “risk of cancer”. Also, important information about the language of causation is lost. Without the disease title, it is unclear to what the causation words are referring. Behind the veil of “[condition]” could be any number of different physical states, each with unique genetic involvement. The audience can no longer assess for themselves how appropriate certain causation language is when combined with a specific condition.

At its extreme, the requirement for anonymity occasionally rendered data unusable. Without specifics, some conclusions could not be effectively illustrated. This fact was most obvious in the discussion of the language of phenotype. It was noted that conditions are often referred to as being "characterized by" a phenotype:

> *Affected individuals were characterized by* [list of characteristics] ...
>
> - Project 2, Paper 2, Line 114

> *Individuals affected by the abnormality were characterized by* [list of characteristics] ...
>
> - Project 2, Paper 3, Line 40
> -

However, the lists of features that were said to "characterize" the condition were different in each of these descriptions. This observation is interesting, but it was impossible to present this data anonymously. The condition being studied, and hence the researchers, could be identified by its characteristics. Therefore, while note was made of the observation, the audience is forced to either accept it on faith or to ignore it. Without quotations, the validity of the analysis is obscured.

The presented language is also impacted by the words and markings that are added to quotations. Through the various place-holders, necessary to maintain anonymity, I have put words into the mouths of the researchers. It is not just a matter of losing information and sense, but also of inserting ideas that were not present initially.

I spoke briefly about my decision to use "condition" as a general place-holder for all condition titles. It was a complicated and trying choice, based on considerations of the meaning of the word and its seemingly innocuous nature. It did not seem to have the negative overtones of a word like "disease". However, as a result of this decision, I ended up putting my beliefs and understandings into my data in a way I had never intended.

The bias present in the repetitive use of "condition" continues to bother me. I was aware at the outset that I would shape my results; that I would construct themes based on my understandings and social context. The presentation of data, through quotations, was also understood to be subjective, in that I was able to choose what and when to quote. These subjectivities are well recognized and expected in qualitative research. I am concerned, however, that the bias that I have inserted into quotations is more concealed.

Quotations allow for the presentation of raw data. The audience relies on this data to judge the adequacy of the author's constructions. In many ways, then, quotations are seen as objective; as the 'truth' in which the analysis in grounded (Strauss and Corbin 1998, Hsieh and Shannon 2005).

I am concerned that the changes made here become so inherent to the raw, 'objective' data that they become a part of it; that the audience, in assessing the data, is actually assessing my words, beliefs and preconceptions. Where I used "condition", others would have used "disease" and the sense of the presented data would have been drastically different.

Nespor (2000) raised a similar, though broader concern in his discussion of the use of pseudonyms for location in qualitative research. He argues that the use of pseudonyms not only strips data of fundamental contextual information, but also reinforces the conception that the data is generalizable. A pseudonym implies that the data is not tied to a specific time and place, but could belong to any time or place. Likewise, my use of "condition" strips away all context and may give the impression that the language used is representative of all situations, not just the specific documents studied.

An example

I have been limited in my ability to present the difficulties of anonymization using my own data, as I cannot present the raw, pre-anonymization data. The best illustration would be the presentation of a sentence from my data before and after changes were made, but this is not possible. Therefore, to provide a more thorough example of the difficulties and consequences of anonymization, I include the following example of a passage from the New York Times. The following two paragraphs are found in a story entitled "Facing Life with a Lethal Gene".

> The gene that will kill Ms. Moser sits on the short arm of everyone's fourth chromosome, where the letters of the genetic alphabet normally repeat C-A-G as many as 35 times in a row. In people who develop Huntington's, however, there are more than 35 repeats.
>
> No one quite knows why this DNA hiccup causes cell death in the brain, leading Huntington's patients to jerk and twitch uncontrollably and

> rendering them progressively unable to walk, talk, think and swallow. But the greater the number of repeats, the earlier symptoms tend to appear and the faster they progress.
> (Harmon 2007)

Of course, a newspaper story from the New York Times is quite different from a set of research documents. In my project I am attempting to hide the identities of the authors: genetics researchers. In a newspaper article, it would be more common to be concerned about the anonymity of the story's subject. For my purpose, I will claim the story was written by a researcher, and that is the identity that I must anonymize.

It is also important to note that the American context of this news story differs drastically from the Canadian context of my research. In the United States the number of researchers studying any given genetic condition far outnumbers those in Canada. Thus, learning that the condition was Huntington's disease might point to hundreds of researchers in the United States, but only a handful in Canada. Such small numbers obviously necessitate a stricter attention to anonymity. Again, for the sake of illustration, let us imagine that the case is in a Canadian setting.

How might we ensure the anonymity of this passage? Obviously, the individual's name and the title "Huntington's" must both be removed. However, it is not clear how they should be replaced. As discussed above, Huntington's could be referred to as a disease, a condition, an illness, a trait, etc. To choose one of these terms requires us to make

assumptions about the author's views of Huntington's. Another option would be to just leave a blank, but that risks affecting the flow and comprehensibility of the passage. Likewise, it is unclear how we should replace the individual's name. One could use a code, such as "Ms. X", although that still identifies her as a woman. Some might suggest "patient", but presently she has no symptoms, just a gene. Does this fit with our, or more importantly the author's understanding of "patient"? Perhaps "carrier" is a better fit. Another option is to use a generic term, such as "person" or "individual", but those terms are far less informative.

The listed symptoms must also be removed, as Huntington's can be easily identified. The status of other medical data is less clear. Knowing that the condition is "in the brain" is certainly identifying, but many conditions occur in the brain. Therefore, in order to decide whether this phrase can be kept, it would have to be compared to all other presented data to assess the additive affect. When combined with other identifiers, such as 'late onset' or 'autosomal dominant', the words "in the brain" may become highly identifying. A choice would have to be made about which words were the most valuable to keep.

The genetics details provided could also identify the condition, but now expertise is required. Without knowing how many genes are present on the short arm of chromosome four, it is impossible to know how distinctive such information is. Similarly, one would have to find out how many conditions are caused by a "repeat of C-A-G" and, of those, how many have a threshold of 35? Without such knowledge, the only appropriate choice

is to remove the data. Could we keep the word “repeat” but eliminate the more specific data? If removed, what can we can we use to replace the highly technical words like “normally repeat C-A-G as many as 35 times in a row”?

The language that originally interested me in this passage was the reference to “the gene that will kill”. However, even this phrase has the potential to identify. Reference to a single gene condition limits the possibilities, and the indication that the gene is lethal further defines the condition. On its own, this piece of information is probably still too vague to pinpoint the authors, but with a few added words – such as “in the brain” or “repeat C-A-G” – it could easily identify. However, as this language is our focus in the passage, its value may outweigh its ability to identify. If used, though, we must be very selective about the remaining presented data.

Ultimately, editing for anonymity leaves us with:

> The gene that will kill [person] sits on [chromosome], where the letters of the genetic alphabet normally repeat [ ]. In people who develop [condition], however, there are more than [number] repeats.
>
> No one quite knows why this DNA hiccup causes cell death in [organ], leading [condition] patients to [list of symptoms]. But the greater the number of repeats, the earlier symptoms tend to appear and the faster they progress.

The resulting passage is awkward and far less compelling. However, several interesting uses of language, including “the gene that will kill” and “DNA hiccup”, are preserved. The data did not have to be completely abandoned – unless, of course, we were interested in the language of repeats, or symptoms, or the name “Huntington’s”.

Assuming that the phrase of interest is “the gene that will kill”, how is the presentation affected by the changes made? It was certainly a more compelling, emotional sentence when we knew the name of the individual destined to die. Without the list of symptoms, death itself is far less interesting. Will this gene allow you to lead a full, rich life and then ‘kill you’ in your eighties? Or will it lead to an early, painful death? We know that the disease can progress more or less quickly, but we no longer know why. Like most of the language presented throughout this study, the general concept seems to have been retained, but the presentation lacks important details and emotional content.

This process is obviously very subjective. Different authors would have made different changes. In fact, if asked to repeat the process, my results would probably differ. Therefore, as much as quotations are used to present ‘the raw data’, they are also the product of the researcher’s personality, preconceptions, and goals. The ability to shape quotations is advantageous; it allows the researcher to select desired data, anonymizing only as necessary. However, it may also allow for inconsistency.

Possible solutions

We have seen that the process of anonymization is difficult, time consuming, and potentially detrimental to the presented data and analysis. Of these, the most problematic to researchers, who spend significant time collecting and analyzing their data, is probably the loss of or damage to results. One possible solution would be to abandon anonymity. This is not an option that can be considered after data collection, especially when anonymity has already been promised. However, if problems with anonymity were anticipated at the outset, participants could be informed prior to consent.

One major advantage of anonymity is the general protection against uncertain harms. In any research, the long term consequences are at best hypothesized. It is difficult to adequately describe all the potential harms. This is especially true in social research, where the harms are more likely to be psychological than physical. When harms cannot be fully predicted, consent may not be truly informed, which is why, even with informed consent, abandoning anonymity is problematic.

In this study, the language in documents was closely analyzed. I was unsure at the outset what I would find, therefore it was difficult to predict what harms might arise. My goal was an exploration of language, but with an ethical perspective I worried that the results would seem more like a critique. As a result, the authors' self-regard and reputation could both be damaged. There is always trepidation in seeing one's work judged. To publish my commentary on language non-anonymously would be akin to posting an old high-

school English assignment, signed and covered in the teacher's red ink for all to see – a scary proposition.

Perhaps if, in addition to informed consent, participants also had the chance to review findings before publication, anonymity could be abandoned. However, such a solution has many problems (risk of researchers being unable to publish, practical limitations, power differential in the researcher-participant relationship) that will be discussed below.

One must also consider the practical implications of abandoning anonymity. Without the safety in being anonymous, research participants may be harder to recruit. Even after being recruited, participants may be less open in non-anonymous studies. In addition to the potential harms, the researcher may want to consider the benefits of maintaining the research relationship, either for the application of study findings or for future research. Finally, it is important to recognize that once anonymity is lost it cannot be regained. Thus, breaking anonymity precludes the researcher from publishing any data that is more sensitive in the future.

An alternative to abandoning anonymity is the member check. By returning to participants and requesting permission to display certain quotations, the researcher provides an opportunity for participants to review what has been written and to choose accordingly (Miles and Huberman 1994). One advantage is that the research participants may be better judges of what counts as sensitive material (Rock 2001). I declined this option on the grounds that it is unfair to the research participants. They entered a research

relationship under the agreement of anonymity and it is unfair to alter that agreement later. Informed consent will also be complicated at this stage, given the relationship between the researcher and research participant, as well as the effort expended by both. A research participant may feel compelled to agree to identification if informed that years of research would be wasted if they did not. Furthermore, if anonymity was promised at the outset, even contacting participants (and hence identifying them) could be unethical.

Furthermore, obtaining secondary consent for identification would have been difficult in this project. Complex documents, like large operating grant applications, often have multiple authors. To obtain truly informed consent, each individual who contributed language would have to be contacted, shown all the results, and then give their consent. The time and effort required by both researcher and participants probably makes such an approach impractical.

The solution employed here was simply to live with the results of anonymity. If one quotation was no longer compelling, others were sought. To remedy lost context, extended descriptions were provided where appropriate. While the presentation was undoubtedly weakened in various ways, I believe that the major emerged themes were still adequately illustrated.

Concluding thoughts on anonymization

The use of quotations is and will continue to be fundamental to the presentation of qualitative data. Quotations are necessary "to provide evidence for some point

(interpretation, claim, or conclusion) the researcher wants to make, to illustrate or provide a more concrete example of an idea, to represent the thoughts, feelings, or moods of the person quoted, to evoke a feeling or mood, or to provoke a response in members of the audience for the research report" (Sandelowski 1994). For document analysis, I would add: to illustrate the potential thoughts, feelings, or responses of the individuals who read the documents.

Chenail (1995) suggests that "data is the star" in the presentation of qualitative research and should be heavily featured. It allows the reader to judge the validity of research and to "see if they can see what you were seeing." It also allowed readers to analyze the data for themselves and invites them "to continue the inquiry and conversation" (Chenail 1995).

However, presenting qualitative data ethically and anonymously can be difficult. The rich and full descriptions strived for by qualitative researchers make identification of research participants easier. In addition to "superficial identifiers", such as names and addresses, qualitative data is full of details that potentially betray identity (Rock 2001). When presenting qualitative data, "the inadvertent description of unique settings and events may reveal the identity of the research participants" (von den Hoonaard 2003).

There are many reasons to support the use of anonymous data in qualitative research. From a sociological perspective, Marx (1999) asserts that anonymity can facilitate communication, encourage reporting of sensitive information, protect from unwanted

intrusions, protect reputation, encourage experimentation, and protect personhood or autonomy. The qualitative researcher, depending on the project's participants, sensitivity, and goals, must consider these implications of anonymity.

In certain situations, identifiability may be permitted or required. The study of previously published data may permit identification. When working with individuals in a position of power, certain information may necessitate, or at least allow the researcher to break anonymity. An example would be the discovery of child abuse. Identifiability can increase accountability and allows longitudinal research (Marx 1999). Finally, some have argued that identification can empower research participants (Rock 2001).

In the opinion of the institutional research ethics board at the University of Western Ontario, this project did not require full research ethics board approval. Therefore, it was not an external obligation that led to the promise of anonymity in our invitation to participate. Although the research ethics board expressed the opinion that this study did not involve human participants, as the focus was documents, I was still cognizant of the potentially sensitive nature of the work. I expected that genomics researchers would be nervous to have information letters and consent forms from ongoing projects analyzed and therefore believed that participation would be higher if anonymity was promised. Furthermore, the analysis of one's writing can be embarrassing and damaging to self-image or reputation. Informed consent could not cover all potential harms, because the results were unknown. I was aware that, despite best intentions, my results could sound

critical. Anonymity offered the participants important protection from harms, even if unforeseen.

The process of anonymization, however, "is full of methodological, ethical, and theoretical tensions" (Thomson, Bzdel et al. 2005) centered on the balance between commitments to research participants and the need to preserve data to produce knowledge (Rock 2001). In the context of archiving qualitative data, Corti (2000) has recognized that "it can be difficult to disguise the identity of participants without introducing an unacceptable distortion into the data." It is likely that "researchers intent on true, complete anonymization rather than formal anonymization face a large, complex task" (Rock 2001).

Data altered to protect anonymity, even if still capable of illustrating the author's ideas, will invariably lose context, content, and complexion. While important for protection of research participants, "anonymizing does change data, so researchers should be aware of both under- and over-anonymizing" (Rock 2001).

Altering data for anonymity also introduces bias into the data. The use of quotations is always subjective, as the "researcher must decide whether to quote, what to quote, and where a quote begins and ends for a particular purpose" (Sandelowski 1994). However, we are apt to see quoted data as objective, direct representations, that can then be judged by the audience. However, the process of anonymization changes those quotes. The researcher places himself in the data, but in a way that may not be obvious to the audience.

Researchers have an obligation to their research participants that ordinarily necessitates anonymity. However, there also exist professional obligations, and pressures, to conduct valid, profitable research and to publish (Homan 1991). Ultimately, though, our obligations to the individuals who participate in our research are paramount.

THEME II. COMPLEX, VAGUE, AND VARIABLE LANGUAGE

The original goals of this analysis were based on an assumption about language use that, after completing the analysis, seems to have been erroneous. The plan was to study the context and usage of genetics related words in order to gain insight into the meanings ascribed to those words by the geneticist authors. I wanted to know, for example, what words were used to describe genetic causation, and how those words varied in different contexts of use. I wanted to know the differences, in the eyes of the authors, between a "disease", a "disorder", and a "condition". I hoped to describe how the authors used words like "normal" and "abnormal".

The primary goal of this analysis was to consider how the definitions of these words within genetics compared to those used more broadly, considering the unconventional predictive and probabilistic way that genetics frames health, disease, and normalcy. I wanted to know how the meanings and understandings of these words might affect communication in general and informed consent specifically. I also hoped to gain a better understanding of the authors' overall views on health, disease, and normalcy through the language that was used.

This section addresses the reasons that I was unable to analyze the language in this way and the erroneous assumptions I had previously made. The title of this theme, "Complex, vague, and variable language", is a simplified statement of the chaos and confusion that I encountered while attempting to analyze this data. Words were intrinsically complex, ambiguous, and seemed to be used interchangeably. I observed this pattern of language

use in many different contexts, including the language used to describe causation, causes, genetic sequences changes, phenotypes, inheritance, disease titles, and the individuals who get sick. This theme focuses on the complexity of the language used, and the resultant potential for miscommunication, by exploring a few of those areas of language.

THEME II. COMPLEX, VAGUE, AND VARIABLE LANGUAGE

SUBTHEME #1: CAUSATION LANGUAGE

The language used to describe the relationship between genome and phenotype was complex. The words and phrases used were familiar, perhaps even mundane. They were remarkable, however, in their abundance and variability. It was unclear how researchers intended these words to be understood, resulting in a bewildering multitude of possible interpretations, whether by myself or by potential research participants.

Direct causation language

The simplest form of causation language, as may have been anticipated, depicts a direct causal relationship between the genome and a particular health condition.

> *Below you will find a brief explanation of our study to identify the cause of* [condition].
>
> - Project 2, Consent Document 1, Line 5

> *A Study to Locate and Isolate the* [condition] *gene*
>
> - Project 2, Consent Document 2, Line 1

> *We wish to find out more about the genetic factors that cause these diseases including identifying some of the genes that are responsible.*
>
> - Project 3, Consent Document 1, Line 15

> [Organization] *previously funded discovery: 1) of causative genes for* [condition] *and* [condition]*; and 2) of >70 human disease mutations causing* [condition], [condition] *and* [condition].
>
> - Project 1, Grant Summary, Line 6

Such language implies that the studied conditions result exclusively from genetics; that to have the gene or the mutation is to have the condition. Direct causation language is simple, easily comprehensible, and fits with some common understandings and representations of genetics (Crosseley 2002, Bubela and Caulfield 2004). Unfortunately, for even the simplest genetic conditions, it is oversimplified and possibly misleading (Lippman 1992, Miller, Begbie et al. 2006). Our understanding of genetics is largely incomplete, and simplistic causal statements likely overstep their scientific basis, even for long-known single-gene conditions (Conrad 1999, Miller, Begbie et al. 2006). Furthermore, this simplistic language reinforces a genetics-centered theory of disease at the expense of other frameworks that include physical, financial, and social environments. The language reinforces a sense of genetic fatalism (Lippman 1992, Conrad 1999).

Direct causation language is, of course, expected and perhaps even appropriate at times. There is no doubt that for certain conditions the presence of a specific allele (version of a gene) or genomic change virtually assures the development of a specific condition within

our current social and physical environment. We can test for alleles and predict outcomes. In these settings, we may wish to refer to the genetic "cause" of a physical condition.[2] When used in informed consent documents, however, direct causation language portrays a potentially misleading, optimistic image of genetic science. The benefits of the study are unintentionally magnified and doubt is displaced. In this way, direct causation language may unintentionally bias the reader in favour of the research and falsely entice participation. Similarly, the implied benefits of direct causation language, when found in grant applications, may unintentionally influence funding decisions.

Susceptibility language

Contrasting the simple cause and effect of direct causation language is the mathematical uncertainty reflected in language that describes causation through probability, susceptibility, predisposition, and risk.

> *These studies will include analysis for genes that affect the risk of developing* [condition].
>
> - Project 7, Consent Document 3, Line 30

> *Research has shown that genetic factors have an effect on the risk of someone developing* [condition] *or other types of* [] *diseases. In other*

[2] The major risk of such language lies in the faulty reasoning that claims that if genetics is "the cause" then it must also be the focus of our medical therapies (Gannett 1999). A "genetic cause" is not external, but rather an integral part of the individual (Shakespeare 1999). In an era where gene therapy remains theoretical, this means that the only genetic therapy available is elimination of the gene – and with it the individual. However, even for the simplest genetic conditions, we know that there are many effective non-genetic treatments.

words, some people are at higher risk of these diseases because of their genetic make-up.

- Project 3, Consent Document 1, Line 11

This permits us to analyze the DNA for the presence of genes, which may confer susceptibility to this disorder.

- Project 2, Consent Document 1, Line 17

This study is designed to identify the genetic (i.e. heritable factors that are passed through your bloodline) risk factors responsible for the susceptibility of [condition].

- Project 4, Consent Document 1, Line 27

Using genetic approaches, we propose to identify major disease genes that predispose to [condition].

- Project 4, Grant Summary, Line 49

Probability is fundamental to understandings of genetics, being inherent to the uncertainty of genetic inheritance, incomplete penetrance, multi-factor causation, and variable gene expression (Hallowell and Richards 1997, Passarge 2007). As is seen in the language used here, the concepts “risk” and “susceptibility” are used to portray this uncertainty.

However, it is important to recognize the inherently negative and therefore potentially pejorative nature of such words (Linell, Adelswrd et al. 2002, Hodgson, Hughes et al. 2005). Direct causation language is relatively neutral, with no intrinsic assumption about the experience of the phenotype.[3] On the other hand, "risk" is an inherently negative concept. It implies that the resulting condition is bad – something to be avoided or perhaps eliminated. (Which is especially problematic in genetics, when conditions are not external, abstract concepts, but rather intrinsic to the individual (Shakespeare 1999).) This language contains inherent assumptions about the importance and impact of the condition. These assumptions are routinely made without input from the impacted individuals. Consequently, primary stakeholders in genetic research describe the concept of "risk" as negative and pejorative (Hodgson, Hughes et al. 2005).

Susceptibility language may also provoke difficulties in comprehension. The concepts of risk and probability are poorly understood, by physicians and the public alike, necessitating the use of clear and simple language in their discussion (Gigerenzer, Gaissmaier et al. 2008). However, the language of susceptibility used here is opaque and largely uninformative. Genes are spoken of as "affecting risk", "conferring susceptibility", or "predisposing to", without reference to the nature or magnitude of that risk. Who would be satisfied with knowing that a procedure entails risk, without being informed whether that risk was 1%, 50%, or 99%; without knowing whether it was a risk of death, impairment, or just minor bruising?

[3] Direct causation language is not completely neutral. The fact that the phenotype has the attention of the medical and research communities, which primarily focus on conditions deemed to be negative, or "disease", does imply negativity. However, this negativity is not inherent to the language. Despite the medical community's negative focus, it is sensible to talk about genes "causing" longevity or happiness. On the other hand, we would never refer to the "risk" of longevity or happiness.

In part, the emptiness of susceptibility language is a result of the emptiness of condition names.[4] Superficially, the nature of the risk is described: it is the "risk of developing [condition]". However, for this statement to be truly informative, the potential research participants must first have an understanding of what that condition entails. Are there varying degrees of severity? Varying definitions? Varying impacts? When these details are hidden behind a disease label, discussions of risk are far less informative.

The lack of detail associated with susceptibility language seen here may be explained by the context of these documents. The research is being conducted precisely to determine the specifics of the genetic risk. The details may simply not be available. Given the prominent position of susceptibility language in genetics, the meanings and understandings of this language should be explored in other contexts. Are descriptions in clinical consent forms richer? Does the language of risk used in conversations about prenatal screening contain information on the nature and magnitude of the risk? If so, how?

Even with clear descriptions, the concepts of risk and susceptibility are difficult to grasp. In one study, only 3% of people even recognized the word "susceptibility" (Erby, Roter et al. 2008). Perceptions of risk differ markedly among all individuals, including health professionals (Kong 1986, Kessler 1990, Hallowell and Richards 1997). Research has shown little relationship between numeric risks and individuals' interpretations. Despite detailed descriptions of probabilities, individuals frequently convert risks into binary equations: either this will or will not happen (Lippman-Hand and Fraser 1979, Parsons

[4] See Section II.4: "Complex, vague and variable language: the phenotype" for further discussion.

1992). Consequently, it is difficult to anticipate how potential research participants will interpret the generalized, superficial susceptibility language seen here. The concepts are theoretically coherent, but may be practically difficult to translate.

As the concept of probability is fundamental to genetics, probability language is likely necessary for clear and accurate communication. Therefore, the difficulties explored here should not be seen as reasons to abandon such language. However, in light of the difficulties in comprehension and potentially pejorative nature of this language, authors should be aware of the implications of these words and strive to make their meanings more transparent.

The distinction between direct and susceptibility language

The distinction between direct causation language and susceptibility language seems clear. The two categories have been recognized previously as distinct, theoretically logical ways to discuss genetic causation (Gannett 1999). As these categories developed, I speculated that the distinction between them would be based on differences between the genetic conditions described. Direct causal language seemed to coincide with common understandings of monogenetic conditions with high penetrance. Susceptibility language, on the other hand, appeared to fit with understandings of more complex conditions involving multiple genes as well as non-genomic determinants. However, like much of the language seen in these documents, the use of causal language was found to be highly variable. Both direct and probability based causation language were used to describe the

same conditions, often in the same document. Consider the language found in project 2's consent documents:

> *Below you will find a brief explanation of our study to identify the cause of* [condition].
>
> - Project 2, Consent Document 1, Line 5

> *This permits us to analyze the DNA for the presence of genes, which may confer susceptibility to this disorder.*
>
> - Project 2, Consent Document 1, Line 17

> A Study to Locate and Isolate the *[condition]* gene
>
> - Project 2, Consent Document 2, Line 1

The information letter begins with the direct and simple "the cause of", but subsequently switches to the obscure "which may confer susceptibility". The title of the consent form then reverts to more direct causation language in its use of "the [condition] gene". The result is confusing. It is unclear how causation is meant to be understood in these documents. Despite the apparent logical divide between direct and susceptibility-based causation language, they are used interchangeably here.

On the other hand, a distinction is drawn between the two types of language in another project:

> *We will now use state-of-the-art* [] *technologies to scan the entire genome in* [] *populations to characterize the full extent and frequency of this newly discovered type of variation, and its potential to cause or influence susceptibility to disease.*
>
> - Project 1, Grant Application, Line 332

Here, the author recognizes and draws attention to the difference between "cause" and "influence susceptibility". In this context, it is clear that the two uses are understood differently. However, no clues are provided as to the basis of that difference.

Thus, it is evident that some authors draw a clear distinction between direct and susceptibility-based causation language, but others either do not recognize this distinction or are careless in their choice of language.[5] Consequently, the descriptions of causation are unclear and confusing. With such conflicting language, the reader's previously formed ideas about genetics are likely to be maintained, regardless of their basis, and consent documents are unlikely to fulfill their role in informing and clarifying. Communication among scientists will also be affected. When some individuals use words synonymously that other individuals view as distinct, miscommunication seems inevitable. The value of words is the shared meanings and understandings they provoke among the individuals who use them, but here these meanings are unclear.

[5] It is possible that two distinct groups of authors were responsible for the language choices. Grant applications are generally authored by researchers, but consent forms are likely to involve input from lawyers or ethicists who may fail to grasp the genetic science being described. An interesting note about consent forms is that they never seem to be signed, leaving the reader in the dark about the influences and potential biases of the author.

Vague causation language: the many remaining words

The third, and collectively largest, category of causation language evolved haphazardly out of an inability to further classify the words. The language certainly did not fit with either direct or susceptibility causation language, but no other features seemed to define it. The words that accumulated within this (supposedly temporary) "other" category were familiar. Some were everyday words like "affect" or "responsible". Others were more technical, like "linked" or "determinant". However, despite countless readings, I could not determine how these words should be grouped or what they could mean.

The lack of obvious meaning connected the many words in this category. Each time I read a sentence, it provoked new thoughts and images. I was lost in trying to determine what these words might have meant to the researchers who wrote them, and what they might mean to the research participants who read them, because I could not even determine what they meant to me. They were just too vague.

Thus, the remaining large category is generally referred to as 'vague causation language'. Although under a single title, the language represented here is not homogenous. Each word provides unique ideas, images, and emotions. However, these words all share a single characteristic that is fundamental to communication and informed consent: they are frustratingly vague.

Of the multitude of vague terminology employed, the most commonly used words were those with the least clear meanings. Genes were said to be “associated with”, “involved in”, or “correlated with” conditions:

> *With the recent increase in understanding of* [genetic process], *it has become increasingly apparent that many mutations and genome variations associated with diseases target the* [genetic process].
>
> - Project 1, Grant Application, Line 945

> *To begin to investigate whether [ ] might delineate unstable regions of the genome, at which new disease-associated rearrangements occur, we will investigate…*
>
> - Project 1, Grant Application, Line 740

> *We may find that some or all of the* [condition]*-associated mutations result in proteins…*
>
> - Project 2, Grant Application, Line 442

> *Further analysis of the* [] *region is necessary to identify the gene involved.*
>
> - Project 2, Paper 2, Line 180

> *Title of Project: Characterization of genes involved in* [organ] *development and* [condition]
>
> - Project 3, Consent Document 2, Line 6

> *… your participation will help us better understand the relationship of* [physiologic parameter]*, and health risks among people to determine correlation with genetic diseases.*
>
> - Project 1, Consent Document 1, Line 178

This language ties phenotype to genome without providing any details about the relationship. How are genes "involved" and to what extent? Are other factors also involved? Is an association even related to causation or does it represent another kind of relationship altogether? This kind of language provides no concrete knowledge about the conditions or genetic processes being discussed.

Why would vague words be used with such frequency? Considering that this language is found in genomics research documents, it might be argued that these words were chosen because of an incomplete understanding of the genetic processes being studied. When we do not know how or to what extent genes are involved, our language is necessarily vague. However, this argument is unconvincing when one notes that the language employed in project 2, a study of an autosomal dominant condition for which the gene was identified decades ago, is the same as the language used in the studies of less well known genetic conditions. Furthermore, the same language is found in both grant applications and

published papers – both before and after research is completed. If a lack of knowledge was a primary force in language choice, finished research would presumably alter word selection. Another possibility is that this language was chosen to avoid technical details that could be confusing for the lay individuals reading the documents. However, the use of the same words in both consent forms and the more technical grant applications seems to contradict this argument.

It is possible that these words are simply so common that the authors fail to recognize them as empty, or that the words were chosen because the authors felt precision about causation was unimportant. We have the option to explicitly define every word that we use, but we generally strive to be precise only about important concepts. However, causation is a central concept in medical genetics, whether one is interested in diagnosis, prognosis, or treatment. When giving informed consent to partake in genetic research, an understanding of causation is essential to understanding the research's potential benefits and harms.

Vague and uninformative language is obviously problematic in the context of informed consent. The purpose of informed consent documents is to inform. It is generally recognized that consent forms must be comprehensible, and many studies have examined mechanisms to improve comprehension (Bjorn, Rossel et al. 1999, Stead, Eadie et al. 2005). Common suggestions include lowering the target grade level, avoiding long and uncommon words, decreasing sentence length, and eliminating jargon (Baker and Taub

1983, LoVerde, Prochazka et al. 1989, Hammerschmidt and Keane 1992, Kimmelman and Levenstadt 2005, Pothier 2005, Farrell, Deuster et al. 2008).

However, the previously unrecognized problem of common but vague language may present a larger obstacle to informed consent than language that is esoteric. Though jargon may not be understood, it is easily recognized, allowing for clarification to be sought. On the other hand, the causation words encountered here are very common, and therefore unlikely to draw attention. Their meanings, however, are anything but clear. It is unclear how such vague but common terminology will be interpreted by readers. Presumably, they will retain previous understandings of genetic causation.[6] Unfortunately, these understandings could vary drastically from those of the authors, resulting in a significant communication gap. Vague language, therefore, may create a substantial obstacle to informed consent because it reinforces prior beliefs rather than fulfilling the role of informing.

Decisions about what and how much information is required to make an informed choice are complicated. There is a delicate balance between providing too little detail and overwhelming the reader with irrelevant technical jargon. However, there is little doubt that the central concepts must be defined.[7] In these documents, which describe research into the health effects of genomic changes, the concept of causation is certainly central. It

[6] This is, of course, only conjecture. To explore how these words are truly understood, it will be important to discuss this language with potential research participants.

[7] One fundamental difficulty with informed consent documents is that they are written for a generalized audience and are developed without feedback. This fails to account for the varying levels of information that may be desired by different individuals, and forces the writer to make assumptions about what the reader will find important.

is impossible to assess the potential benefits of taking part in such research, or its potential risks, without understanding how the genome is connected to phenotype, and, more specifically, how the researcher understands this connection. Vague causation language prevents this insight.

Vague language does not just create problems on informed consent documents. Miscommunication is a potential problem with all of the studied documents. Vague causation language in published papers could be problematic if different scientists subscribe to different causal theories, thus limiting the objectivity of the published literature. More problematic are grant applications, whose intended audiences, much like informed consent documents, often consist of lay individuals. As mentioned above, it is impossible to predict the benefits of genetic research without some understanding of the causal link between genome and phenotype. Important funding decisions could be based on misconception when vague terminology is employed. Vague language, especially when used to describe key concepts, is problematic in any context.

Words like “associated” and “involved” completely evade definition. Other vague causation words hint at the authors’ understandings. In particular, some words were noted to suggest a certain magnitude of genetic involvement, which in general was portrayed as large. In fact, because of the significant implied role of genetics, the words “basis” and “underlie” were originally classified as direct causation language:

Understanding the genetic basis of these conditions will be helpful for developing new treatments and preventative measures.

- Project 1, Consent Document 1, Line 113

The genomic basis of their phenotype will be found using …

- Project 7, Grant Application, Line 134

… discovery of [mutations] *underlying* [condition] *when no mutations were found otherwise.*

- Project 7, Grant Application, Line 163

Identification of functional elements in the human genome is one of the most important aspects of genome analysis, and is a critical requirement for searches for mutations and sequence variants that underlie human phenotypes, since these efforts typically are restricted to the known elements.

- Project 5, Grant Application, Line 3

"Basis" and "underlie" are strong causation words, conjuring the image of a foundation or structural framework. Like the steel girders of a building, here genes are said to underlie phenotypes. Initially, this relationship seems very similar to that of direct causation language, where genes cause phenotypes. However, the imagery of a foundation or structural framework leaves room for variation that was not seen with

direct causation language. While genes may be seen as responsible for major characteristics, like the size, shape, and layout of a building, other influences are implied. Buildings can be painted, decorated, and landscaped differently, and they will be changed by weather and use. Similarly, when genes are said to underlie a phenotype, we can anticipate other forces that will influence the final appearance and experience of the condition.

Despite the clear imagery provoked by these words, they remain open to a multitude of interpretations. It is difficult to know how to apply the mechanistic imagery of foundations and frameworks to the complexity of genetic molecular biology. Are genes more like a blueprint (a common genetic metaphor) or the actual steel girders from which a building is formed? Is each individual girder important, or is it the end product that matters? Human beings, far more so than buildings, are in a constant state of growth, maturation, and change. How does such change affect the imagery of the genetic framework? Furthermore, how are the complex, integral interpersonal interactions that shape all human experiences translated into the imagery of buildings? Although this language provides more clarity than the phrase “associated with”, there remains a vast array of interpretations. The language is vague.

The imagery of a genetic framework does highlight a fundamental question of genetic causation. In the metaphor of the building, we might ask which is more important: the external facade or the purpose for which the building is used? The external facade provides us with our initial impression, and as such is generally the primary concern of

the architect. This is also the primary image produced by the causation language seen here. However, on closer inspection, it is generally the use of the building (a result of social interaction) that is emphasized. After all, structurally identical square rooms could be used as anything from bathrooms, to jail-cells, to museums for fine art.[8] Similarly, it has been extensively argued that it is not the genetic "basis" of an individual, but rather his social interactions that truly define phenotype and disease (Parens and Asch 1999, Shakespeare 1999, Wasserman, Bickenbach et al. 2005). In this way, the imagery provoked by the terms "basis" and "underlie", which focuses our attention primarily on the 'external facade', could be seen as harmful and misleading.

Like "basis" and "underlie", other words were noted that seem to imply involvement of factors outside the genome in the causal relationship. However, words like "contribute" and "determinant" provide little insight into the extent of genetic involvement.

> *The purpose of this study is to understand how inherited and other risk factors contribute to* [conditions].
>
> - Project 7, Consent Document 2, Line 19

> *Our long term goal is to characterize the normal function of the* [] *gene and to determine how mutations within this gene contribute to the pathogenesis of* [condition].
>
> - Project 2, Grant Application, Line 591

[8] We, in fact, see this change in social interaction with buildings all the time. The Louvre was once a palace residence. The Tate Modern was a power plant. While the structures remain the same, different social interactions have turned both into world class museums.

> *OBJECTIVE: To discover and define additional genetic determinants of* [conditions]….
>
> - Project 1, Grant Summary, Line 10

Again, it is unclear how genes "contribute" to conditions, but this language seems to imply the involvement of non-genetic factors. When we are asked to contribute money to a cause, we assume that others will do so as well. We contribute an article or a chapter, but to complete a journal or a book others also must contribute. We may occasionally hear of a "sole contributor", but we recognize this as a special instance, and therefore label it distinctly. Generally, "to contribute" means to be one factor among many.

The language of contribution may, in fact, correlate well with common understandings of genetic causation. For the majority of phenotypic conditions, the genome makes a contribution, sometimes large and sometimes small, that is combined with contributions from other sources. However, occasionally, such as with monogenetic fully penetrant conditions, the genome may be considered a "sole contributor". Other times, the genome may not contribute at all.[9]

Thus, vague language is not necessarily problematic because it is incorrect. The word "contribute" may accurately, though simplistically, describe common understandings of genetics. The concern is that no distinction is made between large and small

[9] This image is somewhat oversimplified. Even if a philanthropist is the sole contributor of funds to a cause, the individuals organizing the event, recruiting the philanthropist, and ultimately utilizing the funds must also be considered. The same is true of genomics. Even with 'purely genetic' conditions, the environment is always present, and therefore will always play a role. Similarly, it is over-simplistic to rule out the genome entirely, as it is present in all phenotypes.

contributions, and that, although implied, the other contributors are frequently ignored. The reader is therefore left to guess at the intended meaning of the author, resulting in the potential for miscommunication.

The possibility of a language gap, in which authors and audience define words differently, was particularly evident with the use of the word “linked”.

> *Mutations in* [gene] *were recently linked with one form of* [condition].
>
> - Project 2, Grant Application, Line 130

> *Earlier studies have suggested that some genetic differences in special populations are linked to these conditions*
>
> - Project 1, Consent Document 1, Line 111

In genetic parlance, a physical condition is “linked” to an area of the genome through a methodology called linkage analysis. It refers to a specific relationship between genetic material and physical condition, albeit not a causal one (Passarge 2007). When considering this technical definition alone, it is difficult to classify “linked” as a vague causation word. However, without knowledge of this technical definition, “linked” lacks clear meaning. It is as empty as “associated” or “involved”, and has countless possible interpretations.

The use of the word "linked" illustrates the difficulties that can arise when scientific jargon overlaps with everyday language. To the scientific community, and the authors of these documents, "linked" has a very specific meaning.[10] To the general public, it could mean nearly anything. The result is a gap in communication that obviously weakens the basis for informed consent, and communication in general.

"Linked" is the only causation term encountered in these documents that I know to have a specific genetic meaning. However, it is possible that each of the words described here has a technical meaning when used by geneticists. If true, scientific communication would be more precise, but the gap between geneticist and lay person would grow substantially wider. These words would be much more difficult to identify as jargon, and traditional filters of jargon, such as editors and research ethics boards, would be faced with an impossible task.

In addition to inhibiting comprehension, the indefinite meaning of vague causation words may also allow interpretations to extend beyond the relationship between genome and phenotype. For example, the word "responsible" could impact views on social and personal responsibility in health.

> *We wish to find out more about the genetic factors that cause these diseases including identifying some of the genes that are responsible.*
>
> - Project 3, Consent Document 1, Line 15

[10] Of course, it is unclear if the authors intended the technical meaning of the word "linked". Presumably, at times they use the word in the same way the rest of us do. When the same word has both a common and a technical meaning, there is no way to determine which usage is intended.

> *This study is designed to identify the genetic (i.e. heritable factors that are passed through your bloodline) risk factors responsible for the susceptibility of* [condition].
>
> - Project 4, Consent Document 1, Line 27

As with all vague causation language, it is unclear how "responsible" is to be understood biologically. The idea that genes are responsible, though, has many potential social meanings. When genes adopt a role of responsibility, our own health obligations could be minimized. In a series of focus groups studying health promotion, Crossley (2002) found that deterministic perceptions of genetics can undermine personal responsibility; "if there's something in your genes there's not a lot you can do." Elsewhere, it has been noted that we often use words that abdicate our responsibility for illness, such as "fall ill", "catch a cold", or "come down with" (Fleischman 1999). Many authors have suggested that a genetic framework is being used more frequently to explain all health issues (Lippman 1991, Lippman 1992, Conrad 1999). The use of the word "responsible" to describe genetic causation may reinforce, even if unintentionally, such genetic determinism.

Although "genetic responsibility" could shift the responsibility away from the individual, the opposite is also possible. Genes are an inherent part of the individual. If genes are said to be responsible, it may be difficult to separate this responsibility from the individual. Thus, "genetic responsibility" could be translated into individual

responsibility for things such as health care costs, insurance, and the social impacts of health (Rothstein and Billings 1992).

Responsibility is especially complicated when applied to “hereditary” or “familial” genes. It is easy to imagine parents being held responsible for the genetic health of their children. This responsibility could take the form of increased medical costs or social stigma. It could also be borne out legally, with the developing concept of “wrongful birth” lawsuits (Pioro 2008). Indeed, the burden of genetic responsibility is already recognized in women who express an obligation to undergo genetic testing as part of their responsibility to family members, in past, present, and future generations (Hallowell 1999). Thus, although we all understand the word “responsible”, in the context of genetics it could have many different meanings.

As a category, vague causation language was large and quite varied. The words share the characteristic of unclear meaning, but vary in the extent and impact of their ambiguity. Thus, they were discussed here in smaller subsets. The distinctions presented, however, are purely theoretical; they are not based on the manner or context in which the words were used. Similar to the blurred distinction seen between direct and susceptibility causation language, vague causation words were used interchangeably and with no obvious link to context. For example, in project 2, which explores a well-known monogenetic condition, we see 3 subsets of vague causation language in addition to the direct and susceptibility type language that was presented above.

> *In summary, we have shown that* [condition] *is* ***linked*** *to an* [specific] *region on chromosome* [] *in a four-generation family affected with the disease.*
>
> - Project 2, Paper 2, Line 177

> *Further analysis of the* [specific] *region is necessary to identify the gene* ***involved****.*
>
> - Project 2, Paper 2, Line 180

> *Our long term goal is to characterize the normal function of the* [] *gene and to determine how mutations within this gene* ***contribute*** *to the pathogenesis of* [condition].
>
> - Project 2, Grant Application, Line 591

The large variety of language used to describe the causal relationship makes a precise understanding of that relationship nearly impossible. The multiple possible definitions of each word are compounded when the words are all applied to the same concept. The authors' intended meanings are incomprehensible. However, due to the general familiarity of the words used, their ambiguity is easily ignored, creating the potential for miscommunication.

Vague language is not unique to these documents. Everyday language, when examined closely, is rife with unclear terminology. In medicine, to mask concern with patients, language without clear meaning, and therefore without clear emotion, is often used;

"condition" is used instead of "disease", "mass" instead of "cancer" (Fleischman 1999). Likewise, vague language is a popular tool of the medical student. When quizzed, it is easier to respond that there is an "association" between smoking and cancer, than it is to describe the precise statistics and pathways. It has also been noted that ambiguous language is used by doctors for the illusion of protection against malpractice (Bedell, Graboys et al. 2004).

In informed consent, however, the use of vague language is problematic (Liss, Aspevali et al. 2004). The purpose of informed consent is to provide information, and the quality of this information is obviously compromised if the language used is understood differently by researchers and potential research participants. On the other hand, so much of the language that we use is vague, that to ensure precise definitions for all words, consent forms would likely become unbearably long and full of unreadable 'legal-ese'. Therefore, it is important to recognize those concepts that are central to informed consent, and to ensure that they are clearly defined. However, there is no clear mechanism to identify the concepts that should be defined. Therefore, while it is likely possible to predict some important concepts, such as causality in studies of genetic outcomes, I believe that the safest approach would involve feedback from primary stakeholders when developing informed consent documents.

External References

As a consequence of my struggle to understand the vague causation language used in these documents, I noticed a significant difference between informed consent documents

and the other research documents studied. When faced with a truly vague statement, such as ‘gene x is associated with condition y’, an industrious reader of research papers or grant applications is left with some recourse. Although not universally present, and sometimes difficult to follow, these documents provide citations to previous literature. The statement ‘gene x is associated with condition y’ is presumably based in prior research, and to some extent that research is accessible to the reader. Thus, there is a way to determine precisely what is known about the “association”, and therefore, presumably, what is meant by the term.

The informed consent documents studied universally lacked citations. Thus, the reader is not only left guessing how and to what extent a gene and condition are “associated”, but is also left without any knowledge of how this association was determined. An “associated gene” could be based on very solid scientific research, but could also be based purely in theory or expert opinion. These distinctions seem fundamental in the understanding of research for informed choice.

THEME II. COMPLEX, VAGUE, AND VARIABLE LANGUAGE

SUBTHEME #2: THE CAUSATIVE AGENTS[11]

In the preceding discussion about the language of genetic causation, three relatively simple, but complexly overlapping categories were presented. However, in presenting the data this way, much of the overwhelming complexity that was encountered in the coding process was obscured. Each causation verb, complex in itself, was modified by surrounding language that indicated varying degrees of certainty, as well as a range in the number and type of both cause and effect. As an example of this added complexity, this section explores the descriptions of causative agents.

Potential genomic causative agents are the focus of genetic research, and therefore are the focus of the documents analyzed. As such, I anticipated that the language describing causative agents would be highly developed and precise. However, I found the words used to be complex, variable, and difficult to interpret.

In the simplest cases, the language portrayed a single causative gene.

> *A Study to Locate and Isolate the* [condition] *gene*
>
> - Project 2, Consent Document 2, Line 2

[11] During coding, my working title for this section was "multiple causation genes". However, much of the language encoded in this section was not about "genes", but rather "mutations" or other "genetic factors". Reviewing the working titles throughout the project, I noticed that I frequently used the word "gene" even though the category involved a much greater range of language. I am not sure why. It is possible that I am a product of a generation that values the gene over other medical explanations. It might be the result of my focus on "genetics" in this study, making the word "gene" more noticeable. Or, it could be that the concept "gene" was the most prevalent and obvious in these documents, leading me to code this way.

> *My DNA sample will not be used for any purpose, other than to look for the* [condition] *gene without my further specific written permission.*
>
> - Project 2, Consent Document 2, Line 42

> *It took almost an entire century between the description of* [condition] *... and the identification of a causative gene.*
>
> - Project 2, Grant Application, Line 128

> *Mutations in* [gene] *... have been found in a variant of* [specific condition]. *Mutations in this gene also cause* [another condition] *and* [a third condition]*...*
>
> - Project 2, Paper 2, Line 44

Single gene causation is simple and easy to understand. Logically, the language fits well with a small subset of conditions, usually called monogenetic conditions. It is a form of language that predominates in media descriptions of genetics, which tend to focus on "the obesity gene" or "the cancer gene" (Petersen 2001, Bubela and Caulfield 2004). However, this highly simplified language likely masks the true complexity of genetic causation (Lippman 1992, Conrad 1999).

Descriptions that focus on a single genetic cause are unlikely to be adequate. All genetic influences are modified by a variety of other causes, including both other genetics causes and environmental influences. In avoiding these other causative factors, this language

could misrepresent the nature of causation in the conditions studied. The genetic scientists that author these documents are certainly aware of these other causative factors. I imagine that other factors are not mentioned because they are assumed to always be present – the background conditions for genetic causation. However, it is not clear that individuals who lack training in genetics will recognize the basic assumptions made by this language. Therefore, a gap in understanding could exist between the authors and their audience.

The issue of unmentioned causal factors is highlighted in the case of a single gene causing multiple conditions. For one gene to cause multiple conditions there must be modifying factors. However, the language emphasizes a single gene as the cause and thus masks this underlying complexity.

> *Mutations in* [gene] *... have been found in a variant of* [specific condition]. *Mutations in this gene also cause* [another condition] *and* [a third condition]*...*
>
> - Project 2, Paper 2, Line 44

Even if we ignore the many non-genetic factors that ultimately influence biologic functioning, a single gene may not be so solitary or simple. Within the one "causative gene" will exist a multitude of potentially causative mutations. These many mutations could result in many different phenotypes, some similar but others drastically different.

Thus, within the language of a single gene, there is a multitude of possible causes, presumably recognized by the geneticist, but again potentially unknown to the reader.

Other descriptions of causation refer to multiple causal genes.

> *Characterization of genes involved in …* [condition]
>
> - Project 3, Consent Document 1, Line 4

> *This permits us to analyze the DNA for the presence of genes, which may confer susceptibility to this disorder.*
>
> - Project 2, Consent Document 3, Line 15

> *Using genetic approaches, we propose to identify major disease genes that predispose to* [condition].
>
> - Project 4, Summary, Line 49

> *We would like to study some of the genes that may cause this* [specific] *problem.*
>
> - Project 3, Consent Document 4, Line 6

> *The laboratory will extract DNA from the sample so that we can study the genes that we think may cause your* [specific] *disease.*
>
> - Project 3, Consent Document 1, Line 28

The concept of a condition being caused by multiple genes is logical, and I imagine generally well understood. We all have many genes, and these all interact in biologic processes. However, the concept of multiple causative genes becomes complicated when it is applied to monogenetic conditions.

> *Our strategy to understand the common* [conditions] *is to study rare human monogenic forms, such as* [condition 1]. *CIHR* [] *previously funded discovery: 1) of causative genes for* [condition 1]
>
> - Project 1, Grant Summary, Line 5

This language contains a perplexing contradiction. A monogenetic condition is, by definition, caused by a single gene, but here the author refers to a monogenetic condition with multiple "causative genes". It is unclear how this is supposed to be interpreted. The multiple genes could be found in a single individual. We have already noted that, even in the simplest genetic conditions, there are always modifying factors. Thus, perhaps "monogenetic" is not intended to mean that only one gene is involved, but rather that one gene plays the most important role.

On the other hand, this language could indicate that in any given individual a single gene is considered to be causal, but in different individuals, different genes are the cause. This interpretation complicates much of the other language seen in these documents. "The [condition] gene" is a common phrase, both in these documents and in everyday parlance.

However, if more than one gene causes a condition, it is unclear which gene should be given this title.

The gene is the causal agent most often described in these documents. The emphasis on the gene (with occasional reference to mutations) is not surprising in genetic research, but it may be illogical. The gene is just an area of the genome that is associated with the production of a protein (Passarge 2007). We all have the same genes. It is variations within the gene – mutations or alleles – resulting in different protein products that contribute to our differences. These documents never explicitly define "gene", but depictions of the gene as the primary causative agent seem to indicate that a non-technical definition is being employed. Thus, the word "gene" appears to have more than one meaning – the technical jargon as defined by textbooks and the meaning developed through common use. When everyday words also have technical meanings, there is a substantial risk of miscommunication, and that miscommunication is likely to remain unnoticed.

Mutations are also depicted as causal agents at times in these documents.

> *We have identified the causal mutation* [] *in the* [] *gene within the original family described by* [researcher].
>
> - Project 7, Paper 3, Line 32

By combining [genetic] *analysis with phenotype assessment, we found significant between-mutation variability for* [physiologic parameter].

- Project 7, Grant Application, Line 33

We are investigating the mutations in [gene] *which result in* [condition]

- Project 2, Grant Application, Line 653

[Researcher] *is investigating* [condition's] *substantial inherited component, looking at the ways mutations (DNA sequence changes) affect* [organ] *development and* [] *function.*

- Project 3, Grant summary, Line 9

Three different point mutations were identified, each causing amino acid substitutions in a portion of the protein predicted to interact with the [receptor]. *These mutations likely cause* [condition] *through a haploinsufficiency of the* [gene].

- Project 2, paper 2, Line 149

There is a complex interplay between at least two proteins in which mutations in either one can lead to [condition].

- Project 2, Grant Application, Line 318

“Mutation” is another jargon word that is widely used in general conversation. Consequently there are many related but different possible definitions. It was impossible to determine what meaning was intended in these documents. “Mutation” could refer to any change in a cell’s DNA sequence (as is indicated in the fourth quote), just those changes that have biologic effects, or just changes with presumed negative impacts (Condit, Archer et al. 2002, Condit, Dubriwny et al. 2004). It is unclear if the definition of mutation is limited to the level of DNA, or whether it can also refer to the resulting physical attribute. In these quotations, mutations are said to occur in both genes and proteins, although the fourth quotation contradicts this and limits mutations to DNA. I have informally discussed the definition of “mutation” with a number of geneticists, and received surprisingly different answers.

Interpretations of “mutation” are further complicated by the inherently negative social understandings of the term. In general conversation, a mutation is a negative term. Individuals describe the word as frightening and pejorative (Chapple, Campion et al. 1997, Hodgson, Hughes et al. 2005). Thesaurus associations include words like “freak”, “monster”, and “monstrosity” (Roget's 21st Century Thesaurus 2008). In short, it is a word that can mean very different things to different people.

To complicate interpretation further, there seem to be other forms of genetic variation that are not mutations. However, these other variations were only hinted at, and never explained.

For newly discovered mechanisms of genomic variation, understanding both the molecular mechanism in disease and the phenotypic consequences are essential. Different types of genomic variation will probably have different effects at the level of the phenotype. Thus, careful characterization of phenotypes is essential for defining the impact of genomic variation on disease susceptibility.

- Project 7, Grant Application, Line 7

A carefully selected subgroup of samples and controls will be evaluated using [method] *for the purpose of discovery of* [genetic variation] *underlying* [condition] *when no mutations were found otherwise.*

- Project 7, Grant Application, Line 163

There are many forms of human genomic variation, including single nucleotide and small insertion/deletion polymorphisms, variable numbers of repetitive sequences and structural alterations of chromosomes.

- Project 1, Grant Application, Line 276

They will serve as an international resource for biomedical researchers to address research inquiry into the spectrum of genomic variation in health and disease.

- Project 1, Grant Application, Line 239

Other terms that were occasionally depicted as causative agents in these documents include “protein”, “allele”, and “loci”.

> *These variants can contain entire genes, and in at least 14 instances they overlapped with disease loci.*
>
> - Project 1, Grant Application, Line 332

> *As shown in Table 4, 15 LCVs already identified seem to overlap with known disease loci. Our experiments will allow us to examine whether these LCV polymorphisms in SC are associated in a non-stochastic manner with some specific population or disease alleles.*
>
> - Project 1, Grant Application, Line 332

> *There is a complex interplay of at least two proteins in which mutations in either one can lead to BDA1*.
>
> - Project 2, Grant Application, Line 589

> *Although an animal model with specific* [condition] *mutations has not been described, analysis of the* [genotype] *mouse can provide important insights into how the mutant protein may contribute to the physiology of these two malformations.*
>
> - Project 2, Grant Application, Line 184

These terms represent less familiar technical jargon. No definitions are provided in these documents, and given the confusion around other technical terms, such as “mutation” and “gene”, it is not clear that reference to any dictionary or textbook will adequately reveal the authors’ intended meanings. Without definition, use of these words with potential research participants could be confusing and even distressing. Furthermore, they expand the variety of depicted “causes”, making it difficult to interpret the authors’ understandings of genetic causation.

Finally, there was a set of words used to describe causative agents that were vague and essentially meaningless.

> *OBJECTIVE: To discover and define additional genetic determinants of* [conditions] *by studying rare monogenic human* [conditions] *in Canadian families and communities*
>
> - Project 1, Grant Summary, Line 10

> *The information you give us will be added to our computer files and used in research on heredity and other risk factors for* [condition].
>
> - Project 2, Consent Document 2, Line 35

> *The purpose of this study is to understand how inherited and other risk factors contribute to* [conditions]. *This part of the study allows us to*

> *combine family history information with laboratory studies to identify genetic and other factors that affect* [condition] *risk.*
>
> - Project 2, Consent Document 3, Line 19

> *We wish to find out more about the genetic factors that cause these diseases including identifying some of the genes that are responsible.*
>
> - Project 3, Consent Document 1, Line 15

Some terms, such as “genetic factor” or “genetic determinant”, indicate that the cause is found in genetic material, but give no further indication of the author’s meaning. The phrase “other risk factors” is not even that specific; it could truly mean anything. These complex, poorly defined descriptions of causative agents were coupled with complex, poorly defined causation verbs and, as is described in the subtheme “Phenotypes”, complex, poorly defined consequences. The result is perplexing.

Although the precise definitions of the various causative agents are uncertain, in most individual sentences the purported cause is clear: it was “the gene”, “the mutation”, or “the genetic factor”. However, much like the language of causation, the highly variable use of these terms made interpretations impossible. I will use the language of project 2, a long known, highly penetrant, autosomal dominant monogenetic disorder, as an example.

Some language suggested a single causative gene.

A Study to Locate and Isolate ***the*** **[condition]** ***gene***

- Project 2, Consent Document 2, Line 2

My DNA sample will not be used for any purpose, other than to look for ***the*** **[condition]** ***gene*** *without my further specific written permission.*

- Project 2, Consent Document 2, Line 42

It took almost an entire century between the description of [condition] *...*
and the identification of ***a causative gene****.*

- Project 2, Grant Application, Line 128

Elsewhere, the language reflected the extra complexity of multiple mutations within the same gene.

Mutations in [gene] *... has been found in a variant of* [specific condition]. ***Mutations in this gene also cause*** [another condition] *and* [a third condition]*...*

- Project 2, Paper 2, Line 44

Three different point mutations were identified… These ***mutations likely cause*** [condition] *through ...*

- Project 2, Paper 2, Line 149

Language suggesting multiple causative genes, seemingly contradictory “the [condition] gene”, was also observed.

> *This permits us to analyze the DNA for the presence of* ***genes, which may confer susceptibility to this disorder.***
>
> - Project 2, Consent Document 3, Line 15

> *We have identified a second* [] *locus at … indicating that* [condition] *is* ***genetically heterogeneous****.*
>
> - Project 2, Paper 2, Line 127

> [] *mutations, however, account for only a subset of the* [condition] *cases, suggesting that the trait is* ***genetically heterogeneous*** *…*
>
> - Project 2, Paper 3, Line 51

Finally, there was language in the grant application that acknowledges a range of possible causes.

> *The observation of these characteristics in multiple families suggests that the* ***gene(s) responsible for*** [condition] *may also be expressed in …*
>
> - Project 2, Grant Application, Line 46

> *Identification of the **gene(s) that cause*** [condition] *is therefore necessary to elucidate its cause. Ultimately, knowledge of the **gene(s) that cause*** [condition] *may also help us understand …*
>
> - Project 2, Grant Application, Line 82

Some of the variability in language might be explained by the different types of documents that are represented. The term "gene(s)" was only seen in the grant application, a document written prior to the research, when one might expect language to express uncertainty. Likewise, the term "genetically heterogeneous" was only seen in the published literature, which could represent a new insight from the research or potentially a more sophisticated target audience. However, considering that project 2 represents a monogenetic condition that has been known for decades, it difficult to explain the degree of variation.

The variable language used makes it difficult to anticipate how the potential research participants will interpret the informed consent documents. "The [condition] gene" is the predominant phrase. However, in the description of the process of specimen collection the authors use the word "genes". This inconsistency could be confusing to potential research participants (as it was for me), but it is not clear how it would influence their understanding or ultimate decisions.

The gap in language between informed consent documents and grant application is more concerning. Although "the [condition] gene" dominates the language of the consent

documents, this term never appears in the grant application. The certainty of this strong, simple, single gene is replaced by the uncertainty of "gene(s)". These documents would have been prepared around the same time, so the language difference cannot be explained by different stages of research. The push to simplify language on consent forms, making them more understandable, could be a reason to avoid a term like "gene(s)". However, it is not clear that a single causative gene is any easier to understand than multiple, and the multiple "genes" is actually used elsewhere in the consent form. Another possible explanation for this language gap is that consent documents and research grants have different authors, with lawyers or ethicists playing a role in the language selection of consent documents.

Whatever the reason, there is a difference in the language geneticists use with funding agencies and that which they use with potential research participants. The simplified, single gene causation used on the consent form is similar to the genetic language that has been described in the lay media. This language has been criticized for 'hyping' genetic research and over-representing the scientific basis for a genetic cause (Lippman 1992, White 2001, Bubela and Caulfield 2004). In informed consent documents, this language could overemphasize the potential benefits of participation and therefore impact decisions to participate. The use of different language in different contexts also increases the risk that core concepts, such as genetic causation, will be understood differently by the researcher and the potential research participants.

Of course, I do not believe that scientists of the caliber and reputation of those who voluntarily submitted their consent document to our scrutiny would manipulate their language intentionally. Indeed scientists who thought there may be even a slight chance that such manipulation existed in their document would likely have ignored our request for their participation, rather than contributing their documents to an ethical analysis of language. I believe the scientists always have their patients' qua research participants' best interests in mind. However, it may be difficult for scientists to find the simplified language considering the complex scientific terminology inherent in their research. Beyond the word selection, a lack of appreciation of the social implications of language choice is likely a problem for many scientists and clinicians in all areas of genetics.

The language of causative agents is one of many related subsets of language that combine to describe the authors' understandings of genetics, health, and disease. The language of each of these subsets was complex, vague, and variable. When combined, it was virtually impossible to guess the authors' intended meanings or the audience's potential interpretations of these concepts.

THEME II. COMPLEX, VAGUE, AND VARIABLE LANGUAGE

SUBTHEME #3: THE INTERCHANGEABILITY OF DISEASE LABELS

One of the primary goals of this study was to assess how various genetics and health related words were defined through their usage. I wanted to know, for example, how "disease" differed from "condition" and "disorder". However, it quickly became apparent that developing specific contextual definitions would be an impossible task. The words appeared to be used interchangeably; no distinguishing features could be delineated. In most cases, the synonymous language use was noticeable only through distant comparison. That is, despite describing very similar concepts, different sections of the documents would use different words. Occasionally, however, interchangeable word use would be highlighted within a single sentence or paragraph.

The word "disease" was noted to be used interchangeably with "condition", "disorder", and "malformation".

> *We will also be investigating the link between these specific genetic conditions as mentioned above and* [specific] ***disease - a condition where*** *…*
>
> - Project 1, Consent Document 1, Line 35

> *These can help to study subjects with strong* ***monogenic forms of common diseases****, such as … Individuals with* ***monogenic forms of common conditions*** *serve as models to help understand mechanism.*
>
> - Project 7, Grant Application, Line 79

> *Previous reports of families with* [condition] *have shown that individuals can inherit the* ***disease*** *as an isolated* ***malformation****, or in association with* ***other disorders****.*
>
> - Project 2, Paper 2, Line 119

> *However, to date, only about 1900 of the some 6000 Mendelian genetic* ***diseases****, and many fewer of the complex common* ***diseases****, have been characterized with any certainty at the gene level. Furthermore, for most of the characterized* ***disorders****, only a subset of causative mutations are known.*
>
> - Project 1, Grant Application, Line 188

Sometimes, such as in the first quotation, the substitution is direct and seemingly purposeful. In other cases, such as the second and fourth quotations, different words are used in two adjacent sentences referring to the same phenomenon, creating the impression of a desire to be varied in language use, but without indicating the relationship between the concepts. Nevertheless, it is evident that in these documents "disease", "condition", "disorder", and "malformation" are, at least at times, used interchangeably.

The relationship between "disease" and "phenotype" is slightly more complex. In some language it would appear that they are treated as equivalent.

> *The severity of this phenotype ranges from* [] *… to a mild form of the disease, in which …*
>
> - Project 2, Grant Application, Line 21

However, elsewhere disease is represented as a subset of phenotype.

> *Furthermore, annotation of the full range of human genomic variation provides a starting point for understanding inter-individual differences underlying various phenotypic conditions, including disease states.*
>
> - Project 1, Grant Application, Line 254

Of course, these two usages are not mutually exclusive. The words “dog” and “animal” can be used interchangeably while still recognizing that dog is a subset of animal. However, as the words are never defined, it is difficult to interpret exactly how the authors intended these words to be understood.

The interchangeability of words developed a complex web. Above, “disease” and “disorder” were used as synonyms, but “disorder” itself is used as a synonym for a large variety of other words, including “condition”, “trait”, “characteristic”, “syndrome”, and “problem”.

… described a large family from [place] *with a* **[specific] *disorder*** *and correctly surmised that* ***the trait*** *was transmitted as an autosomal dominant* ***condition***.

- Project 2, Grant Application, Line 8

You are being asked to take part in a research study that will investigate **[specific] *disorders*** *such as … We will also be investigating the link between* ***these specific genetic conditions as mentioned above*** *and* **[specific] *disease***…

- Project 1, Consent Document 1, Line 29

There are >30 such documented ***genomic disorders****, including* [], [] *and* [] ***syndromes***

- Project 1, Grant Application, Line 421

Some of the most commonly reported associated ***disorders*** *include* []. *The observation of these* ***characteristics*** *in multiple families suggests …*

- Project 2, Grant Application, Line 46

You are part of a family with an **[specific] *disorder*** *that may be inherited. We would like to study some of the genes that may cause this* **[specific] *problem.***

- Project 3, Consent Document 4, Line 6

In turn, many of these words are again used as synonyms for other words, or for each other. Figure 3 illustrates the web of equivalency that becomes evident in these documents. The relationships indicated in this diagram are only those for which direct examples of equivalency could be found, and therefore represent only a subset of what seemed to be a much larger web. In the vast majority of cases these terms were used individually, and their equivalency is only hinted at through similar usage.

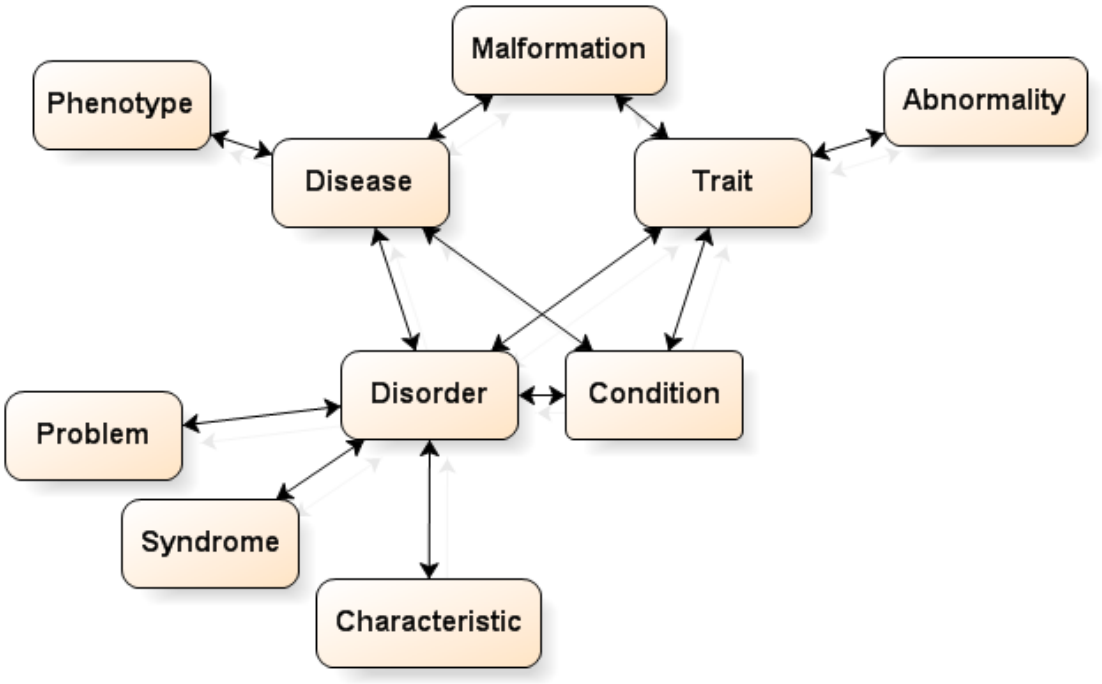

Figure 3. Equivalency of disease words

The highly variable use of such disease titles is well represented in the many names used to describe the condition studied in project 2. Through the various study documents, we see the same condition described as: a phenotype, a disease, a condition, a trait, a genetic disorder, a congenital developmental disorder, an abnormality, and a malformation.

Previous reports of families with [condition] *have shown that individuals can inherit the* ***disease*** *as an isolated* ***malformation****, or in association with other disorders.*

- Project 2, Paper 2, Line 119

Despite being the first ***human trait*** *described in terms of autosomal dominant Mendelian inheritance, little was known about the genetics of this* ***malformation*** *until recently.*

- Project 2, Paper 3, Line 46

The severity of ***this phenotype*** *ranges from … to a mild form of* ***the disease****, in which …*

- Project 2, Grant Application, Line 21

[] *described a large family from* [place] *with a* **[specific]** ***disorder*** *and correctly surmised that* ***the trait*** *was transmitted as an autosomal dominant* ***condition****.*

- Project 2, Grant Application, Line 8

Individuals affected by ***the abnormality*** *were characterized by …*

- Project 2, Paper 3, Line 38

The [conditions] *are a group of rare,* ***congenital developmental disorders*** *that affect …*

- Project 2, Grant Application, Line 2

[Condition] *is a* ***genetic disorder****, which is inherited as an autosomal dominant trait.*

- Project 2, Consent Document 1, Line 8

[Condition] *is a* ***genetic disorder*** *in which …*

- Project 2, Consent Document 1, Line 10

I _____ understand that my/my spouse's family has been diagnosed as having a ***genetic (inherited) condition*** *called* [condition].

- Project 2, Consent Document 2, Line 7

This condition *is quite mild and results in … The inheritance of* ***this condition*** *is called autosomal dominant, which means that …*

- Project 2, Consent Document 2, Line 9

The varied use of disease related words throughout these documents increases the complexity of the language, and increases the chance of confusion or miscommunication. Multiple words are seemingly used to describe the same phenomenon, but it is unclear if

they are meant to be understood as synonyms or just related concepts. The variable use of language makes it difficult to determine exactly what each word refers to. Is “the disorder” the same as “the condition” or “the disease”?

The interchangeable use of “disease” and “phenotype” produces unique difficulties. A phenotype could be understood to represent all aspects of an individual. A disease, on the other hand, will only ever consist of a subset of attributes. So, while I may have the “disease phenotype” for hypertension, other aspects of my phenotype, such as my eye colour or my height, are unaffected. When the words are used interchangeably, the disease becomes the phenotype. An individual’s entire identity is subsumed by the disease title that is applied to them.

Closer inspection of the documents in project 2 indicates that, to some extent, the difficulties of variable language have been recognized. Of the many disease words used, only 2 are found in the information letter and consent form: “condition” and “disorder”. Each word is used multiple times without substitution, indicating a commitment to consistency that is not seen elsewhere. The reason for this isolated consistency is unclear. It could stem from a desire to simplify consent forms and improve readability. The consent forms could have had a single author, while the other documents had multiple. It could have been the result of a research ethics board suggestion to avoid certain terminology. Whatever the reason, there seems to be some effort to limit the variability of language in the documents intended for potential research participants.

Perhaps more important than the limitation of variability is the language selected for use. Three terms that were not used in either the information letter or the consent form were: “congenital developmental disorder”, “malformation”, and “abnormality”. These words are frightening, and would likely be considered pejorative by some (Hodgson, Hughes et al. 2005). It makes sense to omit them from consent documents. However, if such language is inappropriate on a consent form, what makes it more appropriate to use in other documents? This language is found in published literature (papers 1-3) which is potentially available to research participants. These words are also obviously part of the normal lexicon of the researchers and are treated here as the equivalents of other disease related words. While they may be absent from the formal language of consent documents, it is reasonable to assume that these words could arise in the verbal discussions that are also part of the consent process.

In many cases, despite a negative tone, these words fulfill important technical roles. They may be avoided with patients or research participants because of fears that they may be misunderstood, but potentially remain valuable in other discourse. However, that function is only possible when each word maintains its own specific definition. A “malformation” may be quite different from a “disease”, but when the two words are used interchangeably, as we saw in these documents, that difference is lost and the words lose their value.

THEME II. COMPLEX, VAGUE, AND VARIABLE LANGUAGE

SUBTHEME #4: THE PHENOTYPE

An understanding of genetic science requires knowledge of the consequences of the genome. The technical term used to describe the genetic outcome is the "phenotype". Like other genetics concepts encountered in the studied documents, depictions of the phenotype were complex, allowing for many possible interpretations. In some cases, descriptions of the phenotype were peculiar, differing significantly from my expectations and conventional descriptions of health and disease.

Empty condition labels

The word "phenotype" was used relatively rarely in these documents. Clinical descriptions of the phenotype (lists of characteristics or symptoms) were also uncommon. The vast majority of references to the consequences of genomic influence used the name of a condition to describe the outcome.

> *Below you will find a brief explanation of our study to identify the cause of* [condition].
>
> - Project 2, Consent Document 1, Line 5

> *These studies will include analysis for genes that affect the risk of developing* [condition].
>
> - Project 7, Consent Document 3, Line 30

We propose to determine whether certain genomic regions of either deletion and/or amplification in genomic DNA are associated with risk of [condition].

- Project 7, Grant Application, Line 151

Research has shown that genetic factors have an effect on the risk of someone developing [condition] *or other types of* [] *diseases. In other words, some people are at higher risk of these diseases because of their genetic make-up.*

- Project 3, Consent Document 1, Line 11

Mutations in [gene] *were recently linked with one form of* [condition].

- Project 2, Grant Application, Line 130

OBJECTIVE: To discover and define additional genetic determinants of [conditions]….

- Project 1, Grant Summary, Line 10

This language feels natural. We usually refer to medical conditions by a single name, not by a list of symptoms. However, the use of this medical nomenclature presumes prior knowledge of the condition described. In fact, insofar as these names are specialized medical terms they are probably best thought of as jargon (Farrell, Deuster et al. 2008). Sometimes this is obvious. Doctors would never tell a patient, "you have amyotrophic

lateral sclerosis", and simply assume that the patient understands. However, when using common condition names, it is easy to overlook their technical nature. I imagine many patients have been told that they have "diabetes", "pneumonia", or "strep throat" without any further explanation of the terms.

Disease names are only meaningful insofar as they are given a definition. However, just knowing a definition does not ensure understanding. Science is constantly evolving and definitions of diseases continuously change. Furthermore, a single disease title is often applied to a heterogeneous group of symptoms, or conditions that have a wide spectrum of severity. Fleischmann (1999) recounts her frustration with trying to communicate with other patients diagnosed with "MDS". Even though they all shared the same disease title, every individual had different symptoms, causing confusion and limiting the utility of discussion. Indeed, within these documents, when the symptoms were listed for a given condition, they often varied greatly from one description to another.

The use of disease titles without a corresponding definition therefore creates a significant potential for miscommunication. When disease names are common, they are likely to be recognized, but this does not guarantee a common understanding. In fact, given that the researchers will have spent years, if not decades, studying a condition, it is reasonable to assume that their understandings of that condition will differ significantly from someone with limited exposure, even if, or perhaps especially if that individual has the given condition. Many jargon words, such as condition titles, are used commonly making them difficult to recognize as technical terms (Farrell, Deuster et al. 2008). Although some

jargon is necessary in proving technical descriptions, it is generally recognized that when jargon is used a definition or explanation should follow (Makoul 2001). Patients often have preconceived notions about conditions that are potentially distressing and can differ significantly from physicians' understandings (Chapple, Campion et al. 1997, Bedell, Graboys et al. 2004). The definitional ambiguities disguised by simple disease names may be even more problematic in genetic research, where interpretations of "phenotype" appear complicated and unconventional.

The normalized phenotype

When the concept "phenotype" was discussed, the language was complex, variably applied, and difficult to interpret. My preconception was that "phenotype" is a characteristic of individuals; a description of one person's physical characteristics. What I found was that "phenotype" was applied to a range of concepts, in ways that often surprised me.

Certainly, in some cases "phenotype" was discussed in individuals, which is the usage that I would have expected based on previous conversations about health and disease.

> *In addition, we will characterize the phenotypes of normal and diseased subjects…*
>
> - Project 1, Grant Application, Line 1359

Individuals affected by the disorder were characterized by …

- Project 2, Grant Application, Line 10

Affected individuals were characterized by [physical characteristics], a hallmark feature of [condition], but lacked traits common in more severe forms, such as …

- Project 2, Paper 2, Line 114

Our kindred had features similar to the families described by [researcher]. *although some differences in phenotype were noted*

- Project 2, Paper 2, Line 117

This language represents my prior understandings of, and what I believe is the most common way to discuss, health and disease. Individuals get sick. Dad has diabetes and Aunt Sue is "affected by" her arthritis. However, the fourth quotation caught my eye. An individual's features (the kindred's) are compared to those of a family. I had trouble interpreting this – it was unclear to me exactly what it meant for a family to have features. However, the language of phenotype in a family was used commonly.

In the early [time period] *two* [nationality] *families with similar characteristics were described..*

- Project 2, Grant Application, Line 14

Many members in their study were [], *a phenotype that was not observed from our family*

- Project 2, Grant Application, Line 14

You are part of a family with an [condition] *that may be inherited.*

- Project 3, Consent Document 4, Line 6

I _____ understand that my/my spouse's family has been diagnosed as having a genetic (inherited) condition called [particular condition.]

- Project 2, Consent Document 2, Line 7

In this paper we report a family with a relatively mild form of [condition].

- Project 2, Paper 2, Line 114

The phenotype (represented here as characteristics or disease) does not, therefore, seem to be an individual attribute. It is something that can be present in a group of people. These statements challenged my interpretations of both "family" and "phenotype".[12] The "phenotype" does not seem to be something that one can point to; it is not her brown eyes, or her height, or her asthma. It is something that can be shared.

A "phenotype" was also described as something that occurs in cells, body parts, and biologic processes.

[12] For a further discussion of the implications of this language on the concept of family, please refer to Theme IV: The genetic family.

While the [condition] *phenotype is largely restricted to the* [body part], *with some exceptions, the* [condition] *phenotype affects a broader range of* [bodily] *elements.*

- Project 2, Paper 1, Line 202

The lack of phenotype in [condition] *beyond that observed in the* [body part] *suggests that …*

- Project 2, Grant Application, Line 81

We will express the [gene] *mutant in model cell lines…and will assess effects on cellular phenotype …*

- Project 1, Grant Summary, Line 35

we will: …5) extend families and evaluate metabolic, coagulation and inflammatory phenotypes of specific mutation types

- Project 7, Grant Application, Line 41

[Cells] *will be profiled, since the early biochemical phenotype of* [condition] *is a* [specific] *state with altered concentrations of* [biochemical].

- Project 7, Grant Application, Line 148

Thus, "phenotype" is used to describe a wide range of physical manifestations, from molecular profiles to features of communities. In some ways, these descriptions seem

contradictory. If the phenotype is in a body part – in my leg – how is it also shared among the members of a family?

My preconception of "phenotype" probably impaired my ability to see how it was being portrayed. I thought of phenotype as something that occurs in biologic entities, and I could make this understanding fit with the language seen thus far. The phenotype was in the cell, in the body part, or in the person. Even a family could be interpreted, in some way, as a biologic entity. However, these documents also contained language that located "phenotype" in a disease, which obviously required a different interpretation.

> *… new early phenotypes identified in* [genetic conditions] *may serve as new, powerful diagnostic tools to assess* [condition] *risk in the general population.*
>
> - Project 7, Grant Application, Line 88

> *The lack of phenotype in* [conditions] *beyond those observed in the* [] *suggests that …*
>
> - Project 2, Paper 1, Line 259

In this context, it seems clear that phenotype is not used to describe concrete physical characteristics, but rather abstract or normalized qualities. The idea of a 'disease phenotype' seems to indicate a theoretical collection of symptoms, rather than specific features of any one individual. This interpretation would also explain how multiple

individuals in a family, all with unique physical characteristics, can in some way share the same phenotype. Indeed, there even appears to be a formalized system for the normalization of human phenotypes:

> *The "Human Phenome Project" detailed a system for discovery and cataloguing of standardized phenotypes, and correlating these with genomic data.*
>
> - Project 7, Grant Application, Line 77

The understanding of "phenotype" that becomes obvious through this language is that of a standardized description; it is a method of abstracting away from the specifics in order to group individuals together. This is an important tool for research. We study "diabetes", not necessarily the individuals with diabetes. However, it is also a concept that can be confusing when re-applied to individuals. Individuals with the 'same disease' can present very differently. One person with diabetes may have near normal blood sugars and only require some advice on diet and exercise. Another person will have much higher levels, not only putting them at future risk but also making them currently symptomatic, requiring a variety of medications, further testing, and possibly hospitalization. These differences, hidden behind a single disease name, can be problematic in informed consent, as individuals expect their physicians to discuss their personal risks and benefits, not those of the "average person" (Walter, Emery et al. 2004).

As every individual has a specific set of symptoms, signs, and modifying factors, it can be difficult to determine which are worthy of being grouped into the "phenotype". This process of normalization is represented in these documents in the language of "hallmarks" and "characterization".

> *Individuals affected by the disorder were characterized by …*
>
> - Project 2, Grant Application, Line 10

> *Although* [physical characteristic] *is the hallmark feature of* [condition], *a number of other features have been reported including…*
>
> - Project 2, Grant Application, Line 37

> *Affected individuals were characterized by* [physical characteristics], *a hallmark feature of* [condition], *but lacked traits common in more severe forms, such as…*
>
> - Project 2, Paper 2, Line 114

Presumably, the set of "hallmark features" corresponds to the phenotype being discussed. A family can have a phenotype because, despite each family member's unique physical attributes, they share a set of specific, hallmark characteristics. These hallmark characteristics will not be identical in every individual. In a 'family with high cholesterol' every member will have a different blood cholesterol level and unique attributes that modify the impacts of high cholesterol. However, for the sake of communication, we

group, label, and normalize the individual experiences. Thus, although the concept of a familial phenotype was difficult to grasp, it makes sense if we understand phenotype to be an abstract, normalized concept rather than a feature of any one individual.
Language that describes individuals as being "characterized" by physical attributes is problematic. This language allows the author to decide what attributes of an individual are important. In many cases, the individual will likely disagree. A person may be blind, but could rightly object to being 'characterized by blindness'. This language allows individual traits, no matter how trivial to a person's life, to stand in the place of an entire individuality. As such, this language is likely to be seen as pejorative (Parens and Asch 1999, Taylor and Mykitiuk 2001, Hodgson, Hughes et al. 2005).

The other problem with the normalization of "phenotype" is that abstraction away from the specifics of the individual, while helpful for research, can be counter-productive clinically. Grouping together all patients with depression, for example, will mask many different reasons for being depressed and the accordingly different appropriate treatment modalities. Unlike research, clinical work requires a focus on the specifics of the individual.

These two different descriptions of illness – the abstract normalized account and the account of the specifics of an individual – are not mutually exclusive. However, I am not sure that they are always distinguished by our language. I feel that my confusion about the depiction of phenotype in these documents was a result of my being stuck on the specifics, while the documents described the abstract. It took a long time and a lot of

focus to notice the divergence because the language of disease titles, characteristics, and phenotypes can be used to describe both. The usage observed in these documents does not seem to be the same usage we employ in our daily conversations about health and illness. Even among geneticists there appears to be confusion about how "phenotype" should be understood, as some geneticists believe individuals can exist who have no phenotype, but others argue that would be a contradiction (Crusio 2002). Thus, the phenotype could represent a concept that, due to its multiple possible interpretations, could weaken informed consent through unrecognized miscommunication.

New classes of illness

The normalization of "phenotype" removes the concept from the realm of everyday experience. "Phenotype" becomes the property of science; geneticists control which characteristics are worth our attention. However, the resulting constellation of symptoms or characteristics remains generally recognizable as a traditional medical condition. Other descriptions of "disease" or "phenotype" seemed to hint at entirely new classes of illness. In these descriptions, it appears that people can be sick without having any symptoms. The geneticists themselves may not know what the symptoms are; they just know that there is a disease.

> *Recent studies have shown that this disease often affected multiple family members some of whom may not have any symptoms*
>
> - Project 4, Consent Document 2, Line 30

> *We will: … determine the phenotype in pre-symptomatic FPLD children.*
>
> - Project 1, Grant Summary, Line 39

> *Sensitive new phenomic tools can reveal previously unseen markers, sometimes called "early" or "intermediate" phenotypes.*
>
> - Project 7, Grant Application, Line 78

This language represents a significant departure from normal understandings of health and illness. In traditional frameworks, patients present with signs and symptoms that define their disease. Patients were in control; they had the power to decide what ailments warranted a visit to the doctor. This new language represents disease affecting individuals even though "they may not have any symptoms". These are phenotypes in "pre-symptomatic" individuals, allowing us to label their disease earlier. The ultimate power in determining who is healthy no longer lies with patients, but is instead controlled by the findings of geneticists, who can discover and label "early" phenotypes.

This power is evident in the usage of the word normal within quotation marks that was occasionally observed.

> *Our profiling work in Activity 5 will include pooled leukocytes from "normal" healthy individuals and we will have established which group of genes is expressed at detectable levels on microarrays from this work, as well as from previous profiling studies.*
>
> - Project 6, Grant Application, Line 241

> *The goal for profiling* [condition] *in "normal" cell and tissue sources is to generate an atlas of* [] *that is representative of the "normal" adult human population.*
>
> - Project 6, Grant Application, Line 126

> *Recently, in a preliminary study scanning* [] *the genome in 55 'normal' individuals we discovered … that* [unique mutations] *exist in the human genome.*
>
> - Project 1, Grant Summary, Line 39

When I first encountered this usage, I imagined the quotations were being used in a way that I have used them in the past – to indicate skepticism with the general concept that we can normalize human beings. However, the third quotation implies a different interpretation. The researchers found new mutations in these individuals. While they originally believed these individuals did not have a disease phenotype, the finding of a mutation made the researchers questions if these people were indeed "normal". Presumably, now that the mutation is found, the goal is to find the phenotype.

The creation of new classes of illness and the search for new phenotypes is seemingly the result of a gene-centered theory of disease. In the past we might have diagnosed people based on their symptoms, and then went looking for the associated gene. Now, we can diagnose people based on their genes and then search out their symptoms after the fact.

THEME II: COMPLEX, VAGUE, AND VARIABLE LANGUAGE

SUMMARY

The language of these documents was difficult to interpret. My initial goal of developing contextual definitions of key terminology proved to be impossible. Language use was far too complex, ambiguous, and variable to allow inference of the authors' intended meanings.

Despite the complexity in the language presented throughout this theme, my presentation fails to capture the utter confusion I was faced with throughout the process of coding. The organization of language into discrete categories for the purposes of writing produces an artificial sense of order in the presented data. In this analysis, I compare and contrast language usages that, because they exist across many different documents and studies, would not be evident to the audience of any individual document. Furthermore, each category of language described here was modified by countless others, each of which was equally complex. For example, subthemes that developed in conjunction with the language of causation, but were not presented due to limitations on space, included: language describing multiple levels of causation (from gene to protein to organism); language describing who gets sick (from individuals to families to communities); language of certainty and uncertainty (such as 'the gene *may* cause'); language describing heredity; and the lack of language depicting non-genetic causes. The language of each of these subthemes was equally complex, variable, and vague, and therefore the overall depiction of the authors' understandings of genetic causation was chaotic.

Vague or incomprehensible language is problematic for informed consent. The importance of comprehensible consent forms, which eliminate abstruse language and esoteric jargon, is well recognized (Sankar 2004, Stead, Eadie et al. 2005, Beardsley, Jefford et al. 2007). Certainly, unfamiliar or technical language was a factor in my confusion. However, the vast majority of language used was common and superficially uncomplicated. That common language can also confuse was a key insight from this research. Unlike jargon, the equivocal meaning of these common words is unlikely to be recognized. Consequently, the intended meanings of these words will not be clarified, increasing the likelihood of unrecognized miscommunication.

The role of genetics in health and disease is complex and interpreted differently throughout the literature (Lippman 1992, Conrad 1999, Cox and Starzomski 2004, Mykitiuk and Nisker 2010). The genetic framework, based in probability and prediction, varies significantly from traditional depictions of health and disease (Bayertz 1998, ten Have 2003, Bhutta 2004). Furthermore, personal understanding of genetic health conditions will be strongly influenced by individual life experiences (Wittgenstein 1968). Therefore, it is likely that a geneticist, having spent years focused on the potential genetic mechanisms of disease, will have different understandings of genetics related language than the general public. However, the ambiguity and variability of the language encountered in these documents obscures any such differences in understanding.

When definitions are not explicit, readers are forced to guess at intended meanings, allowing other, potentially inappropriate factors to influence understanding. It is

recognized that the media has created significant hype around genetic research that could skew understandings (Petersen 2001, Bubela and Caulfield 2004). The language of these documents subtly reinforces such hype by portraying genetics as the only or major cause, and intimately connecting disease phenotype with genetic changes. Furthermore, these documents are written and distributed by experts and authority figures, and therefore have a sense of objectivity. A bias towards genetics is probably expected in documents arising from genetic research (as any researcher will likely have a bias in favour of his or her own research), but when such language appears in informed consent documents it may inappropriately aggrandize the potential benefits and thus entice participation.

Language may influence informed choice, but again one should not infer that these researchers are manipulating their research participants. These researchers generously and voluntarily participated in this ethics study. They are highly respected clinicians with their patients' and research participants' best interests at heart. I have described new understandings of the ways that familiar language can result in miscommunication and therefore influence informed consent. As far as I know, this is the first time some of these issues have been noted, so the authors of these documents could not have been aware of them. Potential research participants may be enticed into participation, but if they are it would be the result of the unintended and unrecognized ambiguity of the everyday language that was used.

Significant differences were noted between the language used in consent documents and the language found elsewhere in the genetic research documents. In general, the language

in consent documents was limited and simplified when compared to other research documents, such as grant applications and published papers. This is consistent with the push to make informed consent documents comprehensible by writing them in language suitable to specific, lower grade levels (Hochhauser 1997, Baevsky 2008, Jefford and Moore 2008). Unfortunately, biologic processes are inherently complex, and the process of simplifying language also seems to simplify the depiction of those processes. Thus, although the goal of simplifying language is to increase comprehension, the descriptions provided may sacrifice important information and therefore actually inhibit understanding.

Misunderstandings arising from common, but inconsistently interpreted language may also have effects that extend beyond informed consent. Decisions to conduct and fund research may be influenced by language use that emphasizes the importance of the genome. Independent of scientific support, understandings of genetics will also influence medical practice and patient expectations (Liss, Aspevali et al. 2004). Furthermore, conclusions about genetic causality may extend beyond molecular mechanisms and into the social realm, affecting ideas of personal, familial, and social responsibility (Rothstein and Billings 1992, Crosseley 2002). Finally, varying understandings of vague and variable language can contribute to the capacity of our language to cause harm (Rothstein 1991, Nisker and Daar 2006). For these reasons, it will be important to further explore the understandings of the genetics related language discussed here with the authors of these documents and their intended audience.

THEME III. THE LACK OF DEFINITIONS

The process of coding frequently left me confused and frustrated. I was trying to understand words that seemed to defy definition. All that I wanted was a dictionary that could help order the chaos. In retrospect, it seems odd that it took me so long to officially notice that these documents contained no explicit definitions, no glossaries, and no resources to aid those who did not understand.

Of course, the reason for my delayed recognition was the fact that there was nothing to recognize. I was focused on coding, and no coded sentence could reveal what was not there. It is a difficult task to observe the absent. It is also a difficult task to present the absent. As a results section, this could not possibly be shorter. I have no quotations to present.

Once I recognized the lack of definitions, their absence surprised me. I realized, given the nature of the documents, that I would have expected clear definitions. The purpose of informed consent documents is to explain complex scientific concepts to individuals who may have no prior exposure to such concepts. Surely, such a feat requires definitions? Grant applications also seem to warrant definitions. The documents are written to promote new research, often with novel research methods, and certainly with the goal of achieving novel results. I would have assumed that some concepts associated with such cutting-edge work would need to be defined.

After I noticed the general lack of definitions, I re-read the documents specifically looking for any that I may have missed. In a handful of instances explanations of terms were embedded into the text, but these examples were certainly the exception and not the rule.

> *This study is designed to identify the genetic (i.e. heritable factors that are passed through your bloodline) risk factors responsible for the susceptibility of* [condition].
>
> - Project 4, Consent Document 1, Line 27

> [Condition] *is a genetic disorder, which is inherited as an autosomal dominant trait. This means that an affected individual has a 50% chance of passing the disorder to their children.*
>
> - Project 2, Consent Document 1, Line 8

> [Researcher] *is investigating* [condition's] *substantial inherited component, looking at the ways mutations (DNA sequence changes) affect* [organ] *development and* [] *function.*
>
> - Project 3, Grant summary, Line 9

I debated with myself extensively about whether these quotations truly represent definitions. Part of the problem, I have come to realize, is that I was not sure how to define "definition". In each example, a concept is further explained, but the consequence

could be either an “explanation” or a “definition”. The distinction may be important, considering the complexity and opacity of the language seen in these documents. Reflecting on my results, I believe an explanation is any statement that attempts to put a complex concept into plain words, while a definition is a method of formalizing a word’s intended meaning. In other words, when a word is defined, each and every time that word is used it means the same thing. Both concepts are important. Explanations help clarify the complex scientific and genetic terminology – or the jargon – found in genetic research documents. Definitions, on the other hand, can be used delineate even simple concepts to ensure a shared understanding. That is, most of us would not need “health”, “disease”, or “normalcy” explained to us, but we may want to know how the researcher defines these concepts.

I think, therefore, that the above examples are best thought of as explanations. Their presence does indicate that the authors are aware that some of the concepts that they use are complex and difficult to understand. It remains unclear, however, why these explanations are so rare.

When I could not find explicit definitions, I began looking for language or structure that might provide some of the functionality of definitions. I think, in some instances, a reference section might represent a source for further explanation or definition. Readers with enough interest and energy can trace unfamiliar concepts to the source, and in doing so potentially determine their definitions. This process may work for grant applications

and published papers, although I suspect it is rarely used. However, none of the consent documents in this study contained a reference section.

I imagine that the duel processes of written and verbal informed consent partly explain the observed lack of definitions. Researchers probably see their role in the consent process as explaining and clarifying the content of the consent form. In a sense, they function as the dictionary or reference. However, there are problems with such a verbal dictionary. Unlike a static document, there is no guarantee of consistency in the definitions provided (Bjorn, Rossel et al. 1999). Research ethics boards are unable to review this form of information to ensure its adequacy. Furthermore, one of the primary reasons to provide written consent documents is to provide a lasting source of information, as it is unlikely that individuals will absorb all important information from just a single meeting (Ghulam 2006). What is the value of an information letter that does not contain key information?

My biggest concern with relying on verbal clarification of written information is the responsibility it places on the reader. This system forces the reader to identify words that he or she does not understand, and then ask for clarification. In some cases, this might not happen because the reader is embarrassed to admit a lack of knowledge. More importantly, as we have seen in exploring the language of genetic causation, it is possible to not recognize the words that need to be defined. The words may be familiar, but the author and the reader may define them differently.

Putting definitions on the informed consent documents negates this problem. It allows the researcher, with input from a research ethics board and ideally also from representatives of the target population, to identify the key concepts behind the research and explicitly define them from the outset.

THEME VII. THE GENETIC FAMILY

The depiction of "family" that I encountered in these documents seemed unusual. Like all of the language observed, descriptions of "family" were complex, varied, and at times very difficult to interpret. This section focuses on the most consistent usage of "family": the depiction of families as single, coherent, modifiable entities that are defined genetically.

I first noticed that the definition of family employed in these documents might differ from my own when families were described as having characteristics or diseases.

> *In the early* [time] *two* [nationality] *families with similar characteristics were described.*
>
> - Project 2, Grant Application, Line 14

> *This study could lead to a better understanding of how* [condition] *occurs in families.*
>
> - Project 7, Consent Document 3, Line 20

> *I* _____ *understand that my/my spouse's family has been diagnosed as having a genetic (inherited) condition called* [particular condition.]
>
> - Project 2, Consent Document 2, Line 7

This usage of "family" was novel to me. With my preconceived notions about "family" I was unable to grasp the meaning of a 'family with a characteristic'. The image that came to mind was that of a large group of people trying to balance on a bathroom scale, so that the researcher could conclude, "your family weighs one thousand six hundred and twelve pounds". Likewise, a family with a disease indicated to me a group of people with food-poisoning – everyone is sick. Although not explicit in these phrases, it seems obvious that not every member of the family had the disease or characteristic in question. The statement "I understand that my/my spouse's family has been diagnosed as having a genetic (inherited) condition" seems pointless or redundant if every member shares the same disease – who would not already know? Therefore, it appears that these were not statements of observation, but rather of definition.

This idea was reinforced by language that indicated families could possess mutations or genes.

> *We have now shown that two of his families had mutations in* [gene]
>
> - Project 2, Grant Application, Line 259

> *We have also documented this mutation in families from …*
>
> - Project 2, Grant Application, Line 288

> *We have collected DNA samples from 16* [condition] *families and have sequenced the* [] *gene in all except family 14*
>
> - Project 2, Grant Application, Line 401

The concept of a family having a mutation produces seemingly arbitrary exclusions. In the simplest form of genetic heredity, a parent will have a specific allele (what you could call a mutation) and that allele will be passed on, by random chance, to 50% of the genetically related offspring. Therefore, even in the simplest 'atomic family' only half of the members would have a given mutation. Furthermore, spouses are unlikely to be genetically related and so will also lack the mutation, not to mention families that involve adoption, re-marriage, or gamete donation. In the average 'family with a mutation', less than half of the individuals will actually possess the genetic change.

The idea of a family with a characteristic seems even less coherent. At least with a gene, an allele, or a mutation, there may be a single objective entity that some individuals have in common. A mutation, being just a change in a biochemical chain, can probably be identically replicated. I am not sure that the same can be said for most characteristics, no matter how simple. We might say, "Your family has blonde hair", but this is surely an over-simplification. If you compare the actual pigmentation of each individual's hair, you are likely to find tremendous variation – one may be 'platinum blonde' and another 'dirty blonde'.

Therefore, the statement that a family possesses a mutation or a characteristic does not seem to be observational. A family in which all members share the same mutation or characteristic, even if loosely defined, would be exceptionally rare. Instead, these statements are probably best understood to be definitional. They are statements about what it means to be part of a family. This definitional sense is probably best seen in the use of the phrase "disease family".

> *Some* **[familial condition]** ***families*** *with no mutation in a known gene may be sufficiently informative…*
>
> - Project 1, Summary, Line 46

> *Upon identification of the mutation and validation in other* **[condition]** ***families****, our immediate goal will be to …*
>
> - Project 2, Grant Application, Line 571

Membership in a "[condition] family" seems predicated on the presence of the condition, or the mutation in the case of genetic conditions. By combining the family and the condition into a single entity, the meaning of "family" is fundamentally altered and the relationship of the individual members becomes less certain, especially for those who would have not previously considered themselves to have the condition. This may result in individuals perceiving themselves differently, or being perceived differently by family members, based on whether or not they have the mutation or disease.

Imagine a family in which some of the individuals have been shown to have a BRCA2 (breast cancer related gene) mutation. To use the language of these documents, this is a "breast cancer family that has a mutation in the breast cancer gene". It would be difficult to be counted as part of this family without the mutation, but this obviously excludes members of the family we would traditionally recognize.[13] It also potentially provokes emotional and cognitive distress. What are the impacts on a daughter who has not inherited the BRCA2 mutation, when her three sisters and her mother all have breast cancer? People have described mixed reactions to negative genetic testing, with relief balanced by feelings of increased responsibility and concerns of marginalization (Burgess 2001). Might there be a sense of detachment or guilt that is emphasized by our language choice? What about family members who choose not to be tested? We would benefit from research examining how family members who participate in genetics research see themselves in relation to other family members depending on whether or not they are found to carry the "mutation that occurs in the family".

It is clear that the use of "family" that predominates in these documents varies significantly from everyday language. The definition is based on genetics rather than any social relationship. Thus, the definition seems to be based on the ideals and goals of the researchers, to whom the concept of family is valuable insofar as it is associated with

[13] My assumption that families are not normally thought of as "having disease" may reflect my personal bias in language use. Although this language is novel to me, it may be familiar to the potential research participants, who are the intended audience of these documents. I do not belong to a family that has been diagnosed with a 'genetic condition'. I am also medically trained, and thus probably more focused on mechanisms of genetic inheritance and causation that make it unlikely, if not impossible, for every member of a family to share the same genetic makeup. Certainly, families in which multiple members have been diagnosed as having a disease, such as breast cancer, may very well consider their family 'to have' or 'to be affected by' disease. Issues of language use among individual diagnosed with 'genetic conditions' will have to be addressed with such individuals in the future.

genetic disease. This is further evident in the description of how geneticists control the family, expanding and extending the concept based on research goals, and redrawing family trees based on genetics findings.

> *After our extensive clinical characterization**, the family was subsequently extended to include an additional 15 members**, and genetic linkage studies were initiated to identify the gene that causes* [condition].
>
> - Project 2, Grant Application, Line 223

> ***We will … extend families** and evaluate … phenotypes*
>
> - Project 7, Grant Application, Line 58

> ***These families will be expanded** and evidence for causation of the* [ ] *mutations will be obtained ….*
>
> - Project 1, Grant Summary, Line 21

> *Since we discovered a new pod of the family we anticipate a significant increase in this value.*
>
> - Project 2, Grant Application, Line 246

> *We have obtained DNA samples from descendants of* [researcher 1's] *historically relevant family and we demonstrated that this family not only had the same mutation as* [researcher 2's] *families, but that* [researcher 1's] [nationality] *family was also related to the* [researcher 2's] [nationality]

> *family. We have also documented this mutation in families from* [place] *and from* [place] *suggesting that a common founder may be responsible for a significant proportion of* [condition] *cases with* [gene] *mutations.*
>
> - Project 2, Grant Application, Line 286

> *The observation of a common haplotype between the three families suggests that the* [ ] *mutation is an* ***ancestral allele****.*
>
> - Project 2, Paper 3, Line 87

Here we see that the concept "family" is no longer just studied, but fundamentally changed – "extended" for the goals and by the criteria of science. It appears that individuals only count as "family" if they meet the study's criteria. Previously distinct families and individuals, completely unknown to each other, become a single family because of research results or an "ancestral allele". This language leaves little doubt that family is understood as a genetic concept. Genetics has the power to look back in time and discover new 'scientifically true' families.

More than just emphasizing genetics, the language of these documents creates the concept of family that fits the research. This is best seen in the depiction of single person families.

> *The DNA was from family 16, which is currently represented by a single individual.*
>
> - Project 2, Grant Application, Line 419

> *The candidate gene sequencing approach showed that two patients had novel rare missense mutations in* [gene]. *These families will be expanded...*
>
> - Project 1, Summary, Line 20

In each case, a single individual composes a family – though it is clear from the second example that the researchers hope this will not be the case for long. It is difficult to understand what it means for a single individual to be a family. The definition employed here seems to abandon any need for relationship, whether social or genetic. It is evident that the definition of "family" is guided by researcher goals and understandings.[14]

The emphasis on "family" as a genetic rather than a social construct has potential legal and social ramifications (Caulfield 2002). Language that emphasizes "familial disease" and the responsibility of an "ancestral allele" may increase the feelings of guilt associated with passing on a mutation (Cox and Starzomski 2004). This language could also reinforce ideas of legal culpability of previous generations for current health problems (Pioro 2008). The language used here may also strengthen views that 'biologic' families

[14] The idea of letting a single individual stand for an entire family is analogous to the language that was used to describe phenotypes, where single trait was often allowed to represent an entire individual. It is also similar to the language discussed in Theme V (Conceptual meshing) where we see individuals' entire identities tied to their disease state. Obviously, in focusing on a tiny part of the whole, this language drastically oversimplifies the concept it is supposed to represent. The author claims the power to determine what aspect of the whole is essential and ignores the rest.

are in some way more legitimate than families not based on ‘blood or marriage’ and reinforce stigma surrounding adoption, childlessness, and infertility (Holtzman 2008). If we focus on genetic relationships, it is unclear who qualifies as family. Married couples who do not have children are not genetically related – would they count as family? What about families that involve adoption, re-marriage, or gamete donation? Current family law already favours biologic relationships over those based on social ties alone and language of individuals in positions of power, such as respected scientists, that support this definition could have a powerful impact (Deech 1998, Caulfield 2002, Holtzman 2008).

The language used to discuss genetic families produces problems at two extremes. In some cases the individual is allowed to stand for the whole family, while in others the family eliminates any sense of individuality. The use of “condition family” emphasizes the family, provoking conditional membership, and disregarding or disguising the individual members, many of whom will not have the eponymous condition. Even among those with the condition, this language promotes a potentially trivial trait – membership in a “condition family” – to a defining characteristic. At the other extreme, language was used that represented an entire family through a single individual. Again, a potentially trivial attribute of that individual – his genetic makeup, or condition – was used to define the entire family. This synecdochic language, in which either a part is allowed to stand for the whole, or the whole for a part, represents the power of the author to unilaterally assign significance and promote a theory of disease that may not be accepted by those being described (Parens and Asch 1999).

Although I believe that the language used here illustrates the tremendous power that can be associated with language choice, I do not believe that the authors of these documents intended their language to be interpreted in this manner. In fact, given the subtle ways in which language's power is manifested, I imagine that the authors were unaware of the potential social implications of their words. (As some of these social implications are noted here for what I believe to be the first time, it would seem impossible to have been aware of them.) Certainly, training in the medical sciences has not historically included a significant focus on the impacts of written language. These well respected scientists voluntarily contributed their documents to an ethical analysis of their language. These scientists are obviously acting with the best interests of their patients and research participants in mind. That their language could be interpreted as potentially harmful is surprising, and illustrates the complexity of language that can be interpreted differently by different individuals.

The varying meanings and understandings of "family" also impact the communication between researcher and potential research participant. Although a textual analysis, without the input of the authors and audience, can never be certain about a word's intended meanings and ultimate interpretations, it seems that the representations of "family" in these documents differ significantly from the standard understandings of the word. However, the multiple possible interpretations are hidden within a single, seemingly simple and recognizable word, increasing the opportunity for miscommunication. This easily missed gap in language could pose a significant obstacle in the informed consent process.

THEME V. CONCEPTUAL MESHING

The "disease individual"

'Conceptual meshing' is used to refer to a linguistic formulation observed frequently in the studied documents, in which two distinct concepts are combined into a single new entity, usually without clear meaning. It is left to the reader to determine how the components of the distinct concepts are supposed to mesh.

Despite the frequency of its use, conceptual meshing was not noted until late in the coding process. The style was familiar, and in many ways quite natural. It is used commonly in everyday language, and the majority of uses were unremarkable. It was only the cacophonous tone of the phrase "disease individual" that drew attention to the style.

> *For each disease type we will profile* [genetic material] *for up to 30* ***disease individuals*** *such that sufficient quantities of data can be generated ascertain whether there are statistically significant differences in the quantitative distribution of specific* [genetic] *patterns between the* ***disease individuals****, compared to the pooled sample of healthy/normal individuals.*
>
> - Project 6, Grant Application, Line 248

Some remaining subgroup of **[condition]** ***individuals*** *will have* [] *mutations that may be revealed …*

- Project 7, Grant Application, Line 71

We will also examine **[condition]** ***subjects*** *who have no mutations in* [specific genes] ...

- Project 7, Grant Application, Line 98

We will also examine a subset of another 20 **[condition]** ***patients****, for which their mutation has not yet been determined.*

- Project 1, Grant Application, Line 762

We will: i) extend novel **[condition]** ***kindreds*** *with mutations called []; ii) perform [] comparison of* [conditions]; *iii) determine the phenotype in pre-symptomatic* **[condition]** ***children****.*

- Project 1, Grant Sumary, Line 39

The phrase "disease individual" stood out because it was novel and unexpected.[15] It had the appearance of a typographical error, except that it was repeated consistently. Disease was not an adjective used to describe the individual, as in "diseased individual". Nor was it an attribute of an individual, as in "individual with disease". In fact, in this meshed

[15] For the confidentiality of my research participants, [condition] is used in place of the specific condition names being studied. The true tone of these phrases is revealed by substituting a specific condition name for "[condition]", such as "diabetes individual" or "cancer individual". See Theme I: The difficulty with presenting qualitative data ethically and anonymously for further discussion.

formulation, it is not even clear that the individual is given priority. Disease and individual are no longer separate, but a single, new concept: the "disease individual".

As a reader of these documents, I was left to ponder the meaning of this new phrase.[16] I had a reasonable understanding of the word "individual". Interpretation of "disease" was more complex, but I could draw on external references to help clarify meaning. On the other hand, "disease individual" cannot be found in any dictionary, nor even in common usage. The reader is left to rely on personal understandings of "disease" and "individual" to postulate what the combination might mean.

The unique and complex combination was emphasized by the use of more customary language elsewhere in the documents. Individuals were said to be "diseased" or to be "with disease."

> *We will examine individuals with* [condition], *but no mutation in* [specific genes]…
>
> - Project 7, Grant Application, Line 98

[16] One of the important limitations of this study is the inability to comment on how language is understood by other readers. Confusion of the coder could represent usage of language that is generally confusing, but it could also indicate the coder's specific lack of knowledge, an over-analysis of certain ideas, or a cultural and educational background that may differ drastically from that of the intended audience of the documents. An important, and anticipated, next step is to directly explore these ideas with the authors and the intended audience of the documents.

> *The discovery of significant differences in* [gene] *expression in patients with disease, particularly in genes of known medical relevance, may lead to enhanced diagnostic procedures and …*
>
> - Project 6, Grant Application, Line 252

> *In addition, we will characterize the phenotypes of normal and diseased subjects …*
>
> - Project 1, Grant Application, Line 1358

The use of such traditional language highlights the novelty of "disease individual". It seems to imply a specific intended meaning of this new phrase. However, the very similar contexts in which the different usages were found made it difficult to guess at their intended distinction.

Although it was the cacophony of the phrase that originally drew attention to "disease individual", it was the ambiguity that provoked. "Disease" and "individual" are both familiar terms, but their combination is novel. The use of such a compound neologism is problematic for informed choice. The phrase is not defined – either within these documents or in any outside resource. The reader is left to guess at the intended meaning of the author, leaving room for misinterpretation and distress.

When compared to other forms of neologism, conceptual meshing may be particularly prone to miscommunication. A neologism that is completely foreign is likely to be

defined. A conceptually meshed neologism, on the other hand, has a familiar feeling. The individual words are well known, despite the novelty of their combination. The author is therefore more likely to neglect the need for definition, as was seen in these documents with “disease individual”. The reader is also more likely overlook the term, and therefore fail to ask for clarification.

Unlike completely novel words, which could adopt any meaning, interpretation of meshed concepts is presumably limited by the terms involved. It seems clear that “disease individual” has something to do with “disease” and “individual”. However, the possible interpretations are still quite varied. Using the analogy of another meshed concept, ‘fire truck’, we can see that meshed concepts are prone to ambiguity. Among many other possibilities, a ‘fire truck’ could be a truck on fire, a truck painted with fire decals, a truck meant to fight fires, a truck used to start fires, a truck that delivers fire wood, or matches, or lighters, or even a cloud of smoke that takes on the shape of a truck. Further ambiguity arises when one considers the various interpretations of ‘truck’: pick-up truck, dump truck, monster truck, toy truck, eighteen-wheeler, etc… “Disease individual”, being a novel combination of two fairly complex terms, is even more cryptic.

Of course, the phrase ‘fire truck’ is generally well understood to mean a big red vehicle used to fight fires. The term has developed a specific definition based on common usage. “Disease individual”, on the other hand, is a novel combination. Therefore, it cannot have

yet developed a consensus definition. Like the first use of ‘fire truck’, “disease individual” currently requires clarification.[17]

It should be noted that, in the documents studied, the meshed “disease individual” was not seen in consent forms or information letters. It was exclusively found in grant applications. However, even its presence there should raise questions about informed consent. If the phrase is commonly employed by geneticists, it is reasonable to assume that it will eventually be found on documents intended for research participants or patients. More importantly, we must consider the significance of the gap in language between consent forms and other scientific documents. While the meaning of “disease individual” is unclear to me, it is obviously used by the authors to describe a specific concept. The “disease individuals” were, after all, the focus of the research described. The omission of such language from consent forms may represent an inadequate description of the research being conducted. Surely, if the language is important in describing the project to grant committees, it would also be important in the description provided to potential participants. Individuals may rightly want to know that, for the purpose of a study, they are specifically considered “disease individuals”.

The unclear meaning of these terms may also be problematic in scientific communication: both with grant committees and among scientists. It is unclear the extent to which such terms are used and known among geneticists. They were found in four

[17] The context of meshed neologisms is presumably important in determining the need for explicit definition. In literary works, where many such terms arise, and even in day to day life, some degree of ambiguity is tolerable and probably considered artistic. However in scientific literature, and especially on consent forms and information letters, it is problematic.

different projects in this analysis, but the pool of Canadian genetics researchers from which we were drawing was small, so the authors may have known each other, or even shared documents. However, if such terms are not commonly understood, their use in scientific communication could weaken comprehension and jeopardize objectivity.

The phrase "disease individual" is further problematic in its implications for the individual. It is not immediately clear how "disease individuals" differ from other individuals, but the language certainly implies a distinction. (Why else emphasize the "disease"?) In fact, it is not even clear that a "disease individual" should be understood to be an individual. The two concepts are entangled. The disease is integral, so the individuality is presumably diminished (as compared to a 'stand-alone' individual.) Although the presence or absence of a condition may be paramount to geneticists, that same condition may be secondary to the individuals who possess it. Unfortunately, language that meshes disease and individual signals a potentially false and harmful importance of disease.

Language has been recognized as a powerful tool that can shape our thoughts and influence our understandings of social acceptability (Rothstein 1991, Devlieger 1999, Hodgson, Hughes et al. 2005). Numerous studies have indicated a direct impact of language choice on our actions (Zhang and Norman 1995, Zhang 1997, Hall 2006, Heritage, Robison et al. 2007). The ultimate impact of language that meshes the concepts of disease and individuality are presently unknown, but current discussion of the impacts of language may help us understand what such impacts could be.

A significant body of literature has been dedicated to the discussion of the best, or most appropriate, phrases to use when referring to individuals with disabilities (Devlieger 1999, Harris 2000, Byron, Cockshott et al. 2005, Jette 2005, Schuntermann 2005, Jette 2006). Accepted terminology has changed drastically with time, indicating acceptance of the idea that word choice matters (Devlieger 1999). Although there is no consensus on the best language, and accepted usages vary by country, an underlying theme is clear: we strive for language that emphasizes individuality and ensures that disability is recognized as a secondary (at most) characteristic (Shakespeare 1999).

The meshed "disease individual" clearly runs counter to this goal. Rather than recognizing disease as secondary, it becomes an essential, inseparable component. It is not clear how individuality is supposed to be interpreted in this meshed entity. The combination allows a single trait to obliterate a whole identity (Parens and Asch 1999). It also meshes the individual to a concept – "disease" – that has been classified as offensive or pejorative by affected individuals (Hodgson, Hughes et al. 2005).

The analysis of written documents can only reveal potential interpretations of language, rather than the authors' intended meanings. I am certain that the well-respected scientists who generously donated their documents for this ethical analysis would be distressed to see their language interpreted as other than completely encouraging understanding of information and promoting informed choice of their research participants. This analysis highlights the potential problems resulting from the numerous and frequently unintended potential misunderstandings of genetics related language. Any harmful interpretations of

language encountered throughout this thesis are certainly unintentional. Nevertheless, the goal of this research was to highlight and study such unintended harms so that they may be prevented in the future.

Some language, rather than combining individuality and disease, seemed to eliminate the concept of individuality altogether. The individual still exists in the "disease individual". The sense of primacy may be lost, but the individual has not been discounted altogether. Unfortunately, there was one example in the studied documents of individuals being referred to only as a disease:

> *Initially, we will use high-resolution* [technique] *to examine the* [genetic] *composition of the parents of 20* [condition] *having deletions.*
>
> - Project 1, Grant Application, Line 760

This sentence is awkward and difficult to understand. The proposed research aims at examining the genetic composition of the 'parents of 20 *individuals* with genetic deletions'. However, the sentence reads: we are going to study the 'parents of 20 *condition name* having deletions'. The individuals are replaced by their condition name. This phenomenon has been recognized previously, with patients in hospitals often being referred to by a disease name or body part, such as "I am going to see the kidney in room 7" or "I admitted a sickle cell last night". Though not uncommon, such language is unanimously recognized as inappropriate and disrespectful of patients (Fleischman 1999).

There appears to be a spectrum of language that can be used to describe individuals with disease, with "disease individual" being an intermediate between "disease" on its own and "individual with a disease". (Perhaps the far end of this spectrum should be understood to include "individual" with no mention of disease.) The meaning of this spectrum, however, is unclear. It may represent a temporal change in language use, with one form becoming more commonly used. If this is the case, it would presently be impossible to say which direction along the spectrum we are headed. It could also represent a spectrum of power, in which different scenarios call for different tools of language.[18] Finally, it is possible that the various phrases are just the result of arbitrary, unconsidered language choice.[19]

Conceptual meshing creates a further linguistic problem that I have not seen discussed elsewhere in the literature. By combining "disease" and "individual", a new class of words seems to be potentially pejorative or dangerous. Words that are legitimately applied to a single component of a meshed concept may be problematic when applied to the whole. We may want to "eradicate" disease, but this language would be shocking if applied to a "disease individual". [20]

---

[18] Again, the language used here illustrates the power that can be associated with language choice. However, I do not believe that the authors of these documents intended their language to be interpreted in this manner. These well respected scientists voluntarily contributed their documents to an ethical analysis of their language. These scientists are obviously acting with the best interests of their patients and research participants in mind. That their language could be interpreted as potentially harmful is surprising, and illustrates the complexity of language that can be interpreted differently by different individuals.

[19] Variable and seemingly random word choice was consistently seen in other areas of these documents. Please see Theme II: Complex, vague, and variable language for further discussion.

[20] In this respect, the phrase "disease individual" could actually be beneficial in genetics research. A common concern raised about language in genetics is discussion of curing or eliminating disease. While these goals seem laudable, in genetics the condition is part of an individual, and currently the only way to eliminate the disease is to eliminate the person. Thus, substitution of "disease individual" for "disease" may serve as a reminder of the problems with such statements.

Other meshed individuals

Once the phenomenon of conceptual meshing was recognized, it was noted frequently. However, it remained a difficult phenomenon to identify. I have re-read sections of text repeatedly, only to notice an example of conceptual meshing on a subsequent reading. Further exploring the documents and reflecting on everyday language revealed that conceptual meshing is a very common phenomenon, which likely explains why it was often difficult to spot. We use meshed phrases everyday: television dinner, coffee table, picture frame, soup can, bus driver, door man, etc. Its omnipotence suggests that conceptual meshing is an integral component of our language, presumably increasing flexibility and descriptive ability. However, the fact that conceptual meshing is prevalent does not diminish its complexity, especially in novel combinations.

In the documents studied here, "individual" was noted to be meshed with many concepts other than "disease". These other examples of conceptual meshing provide insight into the meaning and impact of the phrase "disease individual". In one case, we see individuals meshed with their place of origin:

> *Furthermore, compared with* [another place] *in which 5 mutations account for >90% of all* [condition] *cases, the diverse spectrum of causative mutation types in* **[place]** ***subjects*** *with* [condition] *creates an excellent resource for evaluating a range of phenotype-genotype variability.*
>
> - Project 7, Grant Application, Line 30

In this section of the grant application, the author is arguing that location is an important aspect of the proposed research. Geography is one of the factors that make this study worth funding. In this context, the characteristic of study participants most important to the researcher/author is their location. The author's values enter the language through conceptual meshing. Like disease, a characteristic that in most settings is entirely trivial is made an integral component of the individual's identity. In fact, the individual as an individual is lost, and a new concept is formed: the "place individual".

Elsewhere, we see individuals meshed with their role as a research participant:

> *As a **research subject**, it is your right to know and understand all the procedures that are to be done in this study*
>
> - Project 1, Consent Document 1, Line 19

> *THE INFORMATION SHEET MUST BE ATTACHED TO THIS CONSENT FORM AND A COPY GIVEN TO THE **RESEARCH SUBJECT***
>
> - Project 3, Consent Document 2, Line 56

> *If you have any questions about your rights as a **research subject**, please call …*
>
> - Project 4, Consent Document 1, Line 81

> *Additionally, we have genotyped 258 patients and 188 **control subjects** …*
>
> - Project 4, Summary, Line 21

These terms, although familiar, are further examples of conceptual meshing. These examples further illustrate how conceptual meshing can take trivial characteristics – such as being involved in a research study – and make them intrinsic to the individual, based on the priorities of the researcher/author. To the researcher, the most important characteristic of a "research subject" is his involvement in the research.[21] Likewise, "control subjects" are important only insofar as they are controls.

I want to re-emphasize that I do not believe that any of the scientists who voluntarily contributed their research documents to this analysis actually think that their research participants are valuable only for their research roles. This language emphasizes the inherent conflict of the medical researcher. They perform their research precisely to help their research participants; they care about their research participants. However, the demands of objective quantitative research are that individuals be considered in the absence of their individuality, as statistics, in order to reduce bias and discover abstract truths. The language used represents that hidden conflict and illustrates the confusion that is possible when the same words can be used in both a specific technical or scientific sense in addition to a meaning more familiar to the potential research participant.

Finally, we have an intriguing example of individuals meshed to another individual.

[21] Of course, the word "subject" in itself is problematic for much the same reason. It is a word that indicates that the primary worth of the individual is his role in research. A subject is abstracted of individual characteristics in order to become a passive object of study.

> *The genotypes were subsequently compared to data from the* **[researcher] *families***.
>
> - Project 2, Paper 3, Line 72

Here, the author meshes individuals with the characteristic that is most important from the standpoint of the paper: the fact that they were previously studied by a specific researcher.[22] In the lives of the individuals mentioned, one can guess that the researcher was no more than a tertiary character. However, in this phrase the very essence of these families is entwined with the research. The researcher is elevated to a role of fundamental significance.

These examples of conceptual meshing illustrate the potential power of such language. The author is able to choose any characteristic, no matter how trivial, and fundamentally entangle it with an individual. These phrases may be descriptive or even logical, but they are based solely on the priorities of the author. The individuals described have no input, despite the fundamentally transformative impacts of the language.

An analogy to consider is of the phrase "street person". This phrase discounts the individuality of the person. The individual, in this context, only matters insofar as he lives on the street. This language allows us to distance ourselves from an individuality that we don't want to consider and focus on an object defined by its lack of a home. Of course, the same individual who at one time is a "street person" may at another time be a

[22] This language also presents individuals as possessions. Like "subject", it paints an image of subservient individuals who are studied, and ultimately owned by the researcher.

"place individual", and at another be a "disease individual", while all the time considering himself "brother", "son", "father", or "person". In each case, it is the author who defines the essential characteristic the person.

The use of "disease individual" is therefore not simply stylistic. It is a tool of language that allows the author to define the purpose of the individual. We can objectify him. We can discuss his disease without having to consider the symptoms and suffering, allowing for 'better' and 'more objective' science. [23]

The X gene and other meshed concepts

In addition to the meshed "disease individual", the concept disease was also noted to be combined with a number of other concepts in these documents, including genes, mutations, loci, and tissue.

> *Using genetic approaches, we propose to identify major* ***disease genes*** *that predispose to* [condition].
>
> - Project 4, Summary, Line 49

> *A Study to Locate and Isolate the* **[condition]** ***gene***
>
> - Project 2, Consent Document 2, Line 1

[23] Again, I wish to emphasize that none of the researchers who generously participated in this study intentionally use language to objectify their research participants. I did not recognize the inherent power of this language until I had been immersed in these documents for nearly two years. Traditional scientific and medical training does not encompass the social powers of language. The purpose of this research is not to identify potentially harms in the past, but rather to explore language so we can be more aware of its impacts in the future.

My DNA sample will not be used for any purpose, other than to look for the **[condition] gene** *without my further specific written permission.*

- Project 2, Consent Document 2, Line 42

The overall goal of the current proposal is to identify major ***disease genes*** *that predispose to the inherited susceptibility of* [condition] *by:*

- Project 4, Grant Summary, Line 30

Human ***disease mutations*** *and SNPs often occur in these sequences…*

- Project 6, Grant Application, Line 4

[A grant] *previously funded discovery… of >70 human* ***disease mutations*** *causing* [condition], [condition] *and* [condition].

- Project 1, Grant Application, Line 6

We may additionally include ***disease tissue*** *for …*

- Project 7, Grant Application, Line 165

These variants can contain entire genes, and in at least 14 instances they overlapped with ***disease loci****.*

- Project 1, Grant Application, Line 331

> *The location of the **disease locus** within the region between* [] *and* [] *was further supported by multi-point linkage analysis (data not shown).*
>
> - Project 2, Paper 2, Line 100

> *While the* **[condition]** ***phenotype*** *is largely restricted to the* [body part], *with some exceptions, the* **[condition]** ***phenotype*** *affects a broader range of* [] *elements.*
>
> - Project 2, Paper 1, Line 202

Like "disease individual" these phrases transform their subject based on the purpose or interest of the authors. In each case a new single concept is formed from two. In some cases, this combination may be usefully informative. In others, the resulting meshed concept may unintentionally confuse, mislead, or harm.

Insofar as conceptual meshing functions to create neologisms, it should be expected and perhaps embraced in research settings. The very nature of research is to generate new ideas, which will not have been named before. Conceptual meshing is a mechanism that can name these new ideas while retaining known terminology. "Disease mutation" may be a perfect example of this usage. Mutations in the human genome are tremendously common and mostly harmless. However, medical research is only concerned with those that cause injury or "disease". Therefore, the meshed concept of "disease mutation" may be helpful and descriptive.

However, the phrase “disease mutation” remains at least somewhat problematic. This phrase still must be recognized as newly coined and therefore without a widely known definition. “Disease mutation” could mean any number of things. Does a “disease mutation” necessarily cause a disease, or can it just be associated with that disease? Does the link between disease and mutation have to be scientifically proven, or can it just be reasonably conjectured? Newly coined, undefined combinations create potential gaps in communication between researcher and research participants, and even among researchers.

More important is the recognition that such phrases represent the beliefs and goals of their authors and should not be taken as objective truths. This point is well illustrated by considering the well-known example of sickle cell disease. Simplified, sickle cell disease occurs when an individual has two copies of the “sickle cell mutation”. When only one mutation is present, the condition is called “sickle cell trait”, and normally does not have negative health impacts, but interestingly is protective against malaria. In this scenario, how are we supposed to understand the phrase “disease mutation”? A single mutation is actually protective against disease. Therefore, some individuals might want to call this a “health mutation”. Others may wish to just not label it at all.

Finally, by meshing the concept of disease to a mutation, we must recognize the potential of our language to cause harm. In describing scenarios like sickle cell disease, such a combination may just be confusing. However, in other situations this language intimately

ties disease to an element of an individual, masking the possibility that such an individual may not see himself as being diseased (Shakespeare 1999).

The conceptually meshed "disease gene" is even more problematic. While "disease gene" is a very common phrase, both in these documents and within the wider scientific and lay literature, it is a perplexing phrase (Conrad and Weinberg 1996). The concept "gene" simply describes a stretch of DNA responsible for encoding a specific RNA molecule (Passarge 2007). Strictly speaking, we all have the same genes. It is our alleles – specific subsets of DNA order within a gene – that vary. It is difficult, therefore, to interpret the meshed "disease gene". The gene is present in every individual, yet there exists no disease that we would attribute to everyone. Therefore, the meshed "disease gene" must take on a meaning that is distinct from the definitions of its individual components.[24]

The meshing of disease and gene is so prevalent in these documents that long before the phenomenon of conceptual meshing was noticed, an entire category was dedicated to exploring the phrase 'the X gene'. Whenever possible, it seems that genes are named after – or meshed with – a disease or trait or physical characteristic. It occurs extensively in scientific and lay literature (Torres 2002). However, the impact and meaning of this language is not entirely clear.

[24] If the definition of "disease gene" is not limited in some way by its components, then it is also fair to assume that "disease individual" could take on definitions that are removed from our common understandings of "disease" and "individual", further complicating our understanding of a "disease individual".

Aside from being conceptually flawed, the combination of a specific disease name with a gene is troubling for other reasons. Most genes are associated with many possible phenotypes, and all phenotypes are dependent on more than just a single gene. What, then, is the significance of a "disease gene"?

As an example, let us consider the well-known 'breast cancer gene'. The first conceptual difficulty is that only a small minority of breast cancer cases are related to the 'breast cancer gene' (Balmana, Diez et al. 2009). The conceptually meshed 'breast cancer gene' seems to imply a more generalized role. Such a combination emphasizes the genetic theory of disease, at the expense of other theories, and may give potential research participants the impression that genetic research is the best, or only means available to investigate cancer. Indeed, it has already been argued that this misconception has substantially influenced funding decisions, explaining the vast sums that have been spent on the Human Genome Project, at the expense of other areas of research (Wilson 2001).

The emphasis on one specific disease may also be troubling. There are actually two well recognized 'breast cancer genes' (National Center for Biotechnology Information 2008). The BRCA2 (short for breast cancer 2, early onset) gene is one of these two genes. There is no doubt that certain alleles of this gene can cause early onset breast cancer. However, the complexity of the term 'breast cancer gene' is evident when one considers that the same gene has also been linked to prostate cancer (Giusti 2003), pancreatic cancer (Murphy 2002), Wilm's tumor (a childhood kidney cancer) (Reid 2005),

medullablastoma (a brain cancer) (Reid 2005, 2008), and Fanconi's anemia (a rare, fatal blood disease) (Alter, Rosenberg et al. 2007).

The phrase "disease gene" makes a strong and potentially misleading statement about the relationship between genetics and disease. The two are fundamentally linked. Genes are made part of disease, and disease part of genes. Many worry that the public's adoption of a genetic framework of disease is oversimplified and extended beyond its scientific basis (Lippman 1992, Conrad 1999, Mykitiuk and Nisker 2010). Although disease is the result of a complex interplay of many factors, such as social and physical environments, economics, personal behaviors, and available health services, genetic explanations are increasingly eclipsing all others (Lippman 1991, Mykitiuk and Nisker 2010). This simplification to and hype of genetic causes has been extensively noted in media representations of genetics (Petersen 2001, White 2001, Bubela and Caulfield 2004). White (2001) argues that geneticists rely on "genohype" to drive their funding and foster these representations through public relations departments and press releases.[25] "Disease gene" is an extreme example of this genetic oversimplification, in which disease and gene are combined into a single, inseparable concept. When adopted by scientists and funding agencies, an oversimplified genetic framework of disease privileges genetic research over all other approaches (Lippman 1991, Wilson 2001, Miller, Begbie et al. 2006). When adopted by physicians and patients, such thinking may compromise care and lead to 'genetic fatalism' (Crosseley 2002).

[25] The language of press releases from genetics studies is an interesting area for future study.

The use of the conceptually meshed "disease gene" in informed consent documents fosters a potentially faulty understanding of health and disease that may over-emphasize the potential benefits of the research being conducted. A "disease gene" causes disease. It is simple and straightforward. It is something that should be studied. By masking the multitude of other contributors and therefore other approaches that may be considered to treatment and research, the language of "disease gene" may be falsely enticing research participants into genetic studies.

Summary of conceptual meshing

Conceptual meshing is ubiquitous in scientific literature and everyday language alike, and is often unremarkable. It is a tool of language that combines distinct concepts into single novel entities. That conceptual meshing is ubiquitous in the English language likely indicates its importance.[26] It provides flexibility, innovation, and control over the ideas we express. Its prevalence also indicates that conceptual meshing is generally comprehensible.

However, the examples of conceptual meshing explored here illustrate the complexity of this linguistic combination. As is the case with "disease individual", the combinations are often novel and open to a variety of interpretations. Familiarity with the individual terms, by both the author and the audience, can mask this novelty and therefore mask the need for explicit definitions. Authors and audience may interpret these phrases differently, creating a gap in language that would weaken informed consent and scientific communication.

[26] It is unclear whether conceptual meshing is also used in other languages.

As we saw with “disease gene” and “disease mutation”, interpretations of meshed phrases can be very complex. In some settings, meshed phrases seem contradictory – such as describing the sickle cell allele as a “disease mutation” when it can actually be protective. Other times, the phrase seems arbitrary, choosing to name a gene after one disease out of a multitude that can be caused. Indeed, as we saw with “disease gene”, the new concept may differ so drastically from the meaning of the components as to cast doubt on any real connection.

Conceptual meshing instills the values and goals of the author into the language. Depending on the needs of the authors, we saw individuals variably meshed with diseases, places, research roles, and other individuals. Similarly, genes, mutations, and tissues could all be intimately linked with a specific disease to emphasize the importance of the research being described.

The impacts of such language are not immediately clear, but are potentially significant. Meshing genetic concepts to disease supports a genetic framework of disease that may be extended beyond its scientific scope. This link could influence research and funding decisions, as well as research participants’ informed choice. Perhaps more concerning are the potential implications of meshing “disease” and “individual”. The phrase “disease individual” diminishes individuality and may be seen as pejorative by the intended audience of these documents.

The true implications of conceptual meshing cannot be assessed in this study, as I have not assessed the intended meanings of authors, nor the actual interpretations of the

readers. I believe the examples used here illustrate a linguistic formulation that, by concealing its novelty, may unintentionally result in confusion when used in informed consent documents. Furthermore, although unintended, the potential for harm exists if the language is interpreted as pejorative by the intended audience of these documents. It is evident to me that the scientists who contributed their documents to this research do not intend to harm or to confuse. However, it is also not clear how these meshed concepts are intended to be understood. The next step will be to explore these phrases with involved stakeholders.

DISCUSSION

There is a potential gap in communication when discussing genetics, health, and disease, in that the same words, when used by different individuals, can have different meanings. Therefore, understanding is not ensured simply through a shared lexicon. Previously, discussion of miscommunication in science and medicine has focused on unfamiliar language or jargon, based on the assumption that the use of 'plain language' would improve understanding (Stableford and Mettger 2007). In essence, this approach attributes misunderstanding to the use of two different languages: 'medical or scientific English' versus 'plain English' (Stead, Eadie et al. 2005). The primary observation of this study was that miscommunication can occur – and may even be more likely to occur – despite everyone using the same language.

Because of its primacy in our lives, we tend to think of language as objective. We assume others will understand what we say, in the way we intend be understood. However, the meanings of words change with time, place, and context (Wittgenstein 1968, Speed 2006). New understandings are neither instantly adopted, nor mutually exclusive. Therefore, different groups will use words differently, with the consequence, or perhaps intention, of not being fully understood by other groups (Holmes 2001, Speed 2006).

Individual experiences, beliefs, and goals shape understanding (Wittgenstein 1968, Holmes 2001). The word "diabetes" will be understood differently by the scientist who has spent decades researching the condition, than by the individual who has spent

decades living with it, than by the person with no prior experience of diabetes. Even within these broad groups, understandings of words and phrases will vary.

It was this vast spectrum of possible interpretations that made the language of these documents so complex. The individual words were not esoteric, unfamiliar, or overly technical. In fact, because the language was so common and apparently simple, difficulties in comprehension were often not appreciated on initial readings. However, upon reflection it became evident that, despite the ease of reading, the authors' thoughts remained unclear; key concepts remained undefined. The language's complexity was initially masked by an aura of familiarity. The very possibility of miscommunication was hidden by the apparent simplicity of the language.

Language that is familiar, but which allows many possible interpretations cannot easily convey new ideas. The author's novel ideas are obscured by the readers' preexisting understandings. Thus, when familiar language is used to describe complex genetics concepts, previously held ideas are unlikely to be changed. Unfortunately, general understanding of medical, scientific, and genetics concepts is poor (Waggoner and Mayo 1995, Bjorn, Rossel et al. 1999, Corrigan 2003, Miller, Begbie et al. 2006). The genetic framework, based on prediction and probability, is complex and non-intuitive (Issa 2002, ten Have 2003, Bhutta 2004). Concepts that are essential to genetics conversations, such as inheritance, susceptibility, and relative risk, are poorly understood by the general public (Erby, Roter et al. 2008). Therefore, the use of such language is likely to create a

gap in understanding. However, this gap might remain unrecognized because the language used to describe these concepts is familiar and appears simple.

This apparent language gap widens in the context of genetics research when, as was commonly seen in the studied documents, common concepts are redefined in unusual or confusing ways in the context of genetics research. Families were described as single coherent entities, capable of having characteristics and diseases. Illness was redefined so that individuals could be diagnosed and labeled despite having no outward indications of disease. Individuals became entwined with concepts such as disease, location, and research role, to create novel and poorly defined entities. The concepts changed, but the words remained the same, resulting in an invisible obstacle to clear communication.

Changing definitions of health, disease, and normalcy in the context of genetics have previously been recognized (Issa 2002, ten Have 2003, Cox and Starzomski 2004, Mykitiuk and Nisker 2010). Genetic testing creates "presymptomatic" or "potential patients", who have genes, but may never become sick (Bayertz 1998, ten Have 2003). Genetics also creates entirely new categories of disease, in which classic symptoms never develop. For example, pharmacogenetics – the field of study that relates genetics to the efficacy of drug therapies – has the potential to create a new, subclinical classification of disease, in which individuals with a genetically decreased response to a medication are seen as ill (Issa 2002). The meanings and understandings of these new, poorly-defined genetics-impacted concepts are often markedly different from those that are still commonly employed outside of genetics (Cox, Burger et al. 2002, Mykitiuk and Nisker

2010). Thus, a significant language gap could easily exist between those working in genetics and the individuals with whom they interact, including patients and research participants.

Beneath the veil of familiarity, the language of the studied documents was complex. Many words and phrases had the potential for multiple interpretations. Neither the authors' intended meanings nor the potential research participants' likely understandings were clear. The primary source of this confusion appeared to be language that was vague and variable.

Vague language is ubiquitous in all communication. The general meaning of a word can be known without knowledge of a precise definition. For example, everyone understands the word "love", but would likely be challenged to provide a specific definition. Other words are just inherently vague, in that they describe relative rather than absolute concepts. In informed consent documents, such ambiguous terminology can lead to miscommunication. The words "rare" and "common" are used to describe frequency, but despite the seemingly simple meanings of these words, interpretations will vary. Rare to one individual will mean ten percent, while another will picture one in a billion (Sutherland, Lockwood et al. 1991, Beardsley, Jefford et al. 2007). Similarly, interpretations of severity words, such as "mild", "moderate", and "severe", vary enormously even though the words are well known (Wertz and Knoppers 2002, Beardsley, Jefford et al. 2007, Burgess 2007). In these documents, the descriptions of causation, causative agents, and phenotypes were superficially straightforward, but

ultimately ambiguous. The words were comprehensible, and therefore unremarkable, but lacked a clear definition.

Vague terminology is presumably unavoidable. As the definition of any word is developed through shared understandings within a group of language users, differing patterns of usage and differing life experiences will lead to varying interpretations of the same word (Wittgenstein 1968). Minor differences in understanding are often unimportant or can be compensated for through the context of usage. For example, although the definition of "car" may or may not encompass vehicles such as minivans and pick-up trucks, the statement "get in the car" is perfectly comprehensible when standing next to a single vehicle with its door open. On the other hand, when making the complex decision about participation in genetics research, the unintentional use of vague terminology in descriptions of core concepts is obviously problematic.

If every word is potentially vague, vague words cannot be blamed for miscommunication. In situations when precision is important, key concepts must be defined. Definitions provide the audience with access to the author's intended meanings, and therefore allow for the identification of potential discrepancies in understanding. However, the studied documents were completely devoid of definitions, despite discussing complex genetic theories that seemed to fundamentally transform understandings of health and disease.

The importance of defining jargon and complex scientific terminology is well recognized (Makoul 2001, Stead, Eadie et al. 2005, Stableford and Mettger 2007). However, the

language that caused confusion in these documents was generally plain and familiar. Plain, but vague language is open to many interpretations, but this ambiguity is easily overlooked because the words are familiar. The data suggests that these familiar words are also unlikely to be defined or explained by the authors, presumably as a consequence of their familiarity. Therefore, each potential research participant will develop his or her own understandings, without recognizing other possible interpretations, resulting in the potential for unrecognized miscommunication.

My analysis also suggests that understanding can be impeded by inconsistent word choice. I had anticipated that many core concepts would not be explicitly defined in these documents, but had assumed that clues to the authors' intended meanings would be present in the context of their use. In other words, I had planned on developing contextual definitions by comparing and contrasting the manner in which words were used. However, words appeared to be used interchangeably, without obvious pattern, making it impossible to guess at the authors' intended meanings.

The exceedingly variable word choice that was encountered evoked the image of an author immersed in a thesaurus. However, the words used are complex, and it is not clear that they truly represent synonyms. What was referred to in one sentence as a "disease" would be, in the next sentence a "disorder", a "malformation", or a "syndrome". The same condition would be caused by a "mutation", a "gene", multiple "genes", "proteins", or "genetic factors". This variability made it difficult to define the individual words, and also obscured the general concept that these words were meant to describe.

In the process of writing up my results, I gained a better understanding of the phenomenon of variable language use. The nature of the data that I was presenting meant that my writing was, in places, very repetitive. I recurrently used the same phrases to introduce quoted material. I also attempted to limit my impact on the presented data and to avoid potentially pejorative terms by remaining consistent in my use of certain genetics-related terms. For example, I elected to use the word "condition" for all descriptions of phenotypes and disease processes. The resulting repetitive language felt elementary, bland, and awkward – especially in a graduate thesis. I was incredibly tempted to vary my word choice. I cringed whenever the same word was used in subsequent sentences. I didn't want my writing to bore. More importantly, I was cognizant of the fact that my work would be judged as a representation of my academic career. As such, I felt obliged to display a considerable lexicon to prove my erudition and demonstrate that I belonged in the academic community.

I do not doubt that varied language use makes writing more interesting, but it comes with a cost. This study suggests that when words are used interchangeably, each word loses meaning. Instead of denoting multiple unique but related ideas, the words all adopt the same ambiguous sense. Words that in some contexts may have specific technical definitions, such as "gene", become blurred with less specific concepts. The result can be confusing and difficult to follow. It is unclear when multiple words are used whether the authors of informed consent documents intend them to be understood as distinct or as referring to the same concept. Is the "disease" mentioned here the same as the "disorder" mentioned there?

Variable word choice probably also reduces the likelihood that clarification of unclear meaning will be sought. When a single word is not understood, we might expect potential research participants to seek clarification. However, I imagine that many individuals would be hesitant to ask for definitions of multiple words, and therefore may not ask at all. Moreover, when a single word is used repeatedly, it is likely to draw the attention of potential research participants and indicate a key concept. When many different words are used to refer to the same concept, it is less likely to be noticed.

The reason for the use of particular words and phrases is a fundamental determinant of their appropriateness. For example, vague and variable language may enhance the overall aesthetic and multiple meanings of a poem. Poets want their readers to emote, speculate, and be inspired. On the other hand, inconsistent and vague language is detrimental to consent documents, whose purpose is to inform, clarify complex concepts, and empower choices.

In addition to the vague and variable language noted above, a few specific language uses were identified that seem to contribute disproportionately to the language gap between researcher and potential research participant. Jargon, or highly technical terminology, has previously been noted as a source of medical miscommunication (Makoul 2001, Farrell, Deuster et al. 2008). However in medicine, and particularly in genetics, there may be a subset of language that has a specific scientific meaning, but also a common, non-technical usage, and therefore might not be recognized as jargon. Words like “gene”, “mutation”, “carrier”, and “linked” all have precise meanings in genetics, but are used

differently in everyday language. Similarly, specific disease titles, such as “diabetes” or “cancer”, have medical definitions that do not necessarily match those of common parlance. These differing definitions are hidden within a single word that remains unchanged no matter what the intended meaning.

The linguistic formulation that I have called “conceptual meshing” may also significantly contribute to the language gap. My data suggests that, unlike traditional neologisms, a conceptually meshed neologism is unlikely to be defined, as each individual word is familiar and therefore the combination does not appear to require definition. However, although the authors of consent documents using this novel combination have an intended meaning, that meaning may be unclear to the reader. Indeed, as frequently occurred with me, the reader is unlikely to recognize the meshed term as novel or requiring clarification. The result is an unintended gap in language and a potential for miscommunication.

Another contributor to the language gap is the difference between the language used in information letters and consent and that used in grant applications. Some specific examples were noted. “Disease individual” was only found in grant applications, never in consent documents. In project 2 we saw “the gene” in the consent form become “gene(s)” in the grant application. Many condition titles, such as “malformation” and “disorder”, were conspicuously absent from documents intended for research participants. However, the difference is far more than these specific examples, and far more difficult to describe.

Essentially, consent documents seemed relatively simple while grant applications caused confusion and provoked anxiety.

It is not surprising that the language of grant applications and consent forms differ. The target audiences are different, and there is great emphasis on writing consent documents in plain and elementary language (Hammerschmidt and Keane 1992, Goldstein, Frasier et al. 1996, Hochhauser 1997, Stableford and Mettger 2007, Baevsky 2008, Jefford and Moore 2008). However, there is a risk in attempting to simplify the language of consent documents: once language has been changed, it may no longer say the same thing. If "disease individual" and "malformation" were important terms in explaining research to funding bodies, might they not also be important in explanations to potential research participants? As each group is only exposed to the documents intended for them, the differences in language are invisible; there is an undetectable language gap.

Impacts on informed consent

Based on my observations, informed consent is fundamentally about communication. Language is used to promote understanding and request consent. Language must therefore be clear and descriptive, as any miscommunication will weaken the foundation of informed consent. Previous literature examining language in informed consent has tended to focus on difficult language or jargon as barriers to understanding (for example: (Waggoner and Sherman 1997, Davis, Holcombe et al. 1998, Bjorn, Rossel et al. 1999, Coyne, Xu et al. 2003, Stead, Eadie et al. 2005)). The language observed here highlights

a different problem in the unrecognized miscommunication that can occur when common words are understood differently by different individuals.

Any impediment to comprehension is problematic for informed consent. However, this analysis has led me to the observation that common but ambiguous language may present a larger obstacle than language that is esoteric or unknown**.** Although scientific jargon may not be understood, the potential research participant at least appreciates that it is unfamiliar, allowing it to be identified and clarification to be sought. On the other hand, the words discussed here were generally common and therefore unlikely to draw attention. Potential research participants are not prompted to seek clarification and the author is also unlikely to recognize the need for definition. This potential language gap is like a game of 'broken telephone' with everyone speaking English, but the message not transmitted, whereas the use of jargon is like a conversation between one individual who speaks only English and another who speaks only French - the miscommunication is clear (Stead, Eadie et al. 2005).

It is unclear how the common but vague language used in these documents will be interpreted by potential research participants. The authors provide no definitions and there are no signals to the potential participants that a novel meaning is intended. Therefore, the potential research participants' prior understandings are likely to be retained. For example, when the authors of these documents discuss "family", the reader will recognize the word and interpret it as always, even though the intended meaning may

be novel. Thus, this language may promote retention of previously held ideas and interfere with informed choice.

In promoting the retention of previously formed ideas, the language gap also allows potentially inappropriate factors to influence understanding. Pre-existing bias and misunderstandings are not addressed. This is especially problematic in genetics research. The field of genetics is extensively discussed in the media, and depictions tend to "hype" genetics, overstating the capabilities of current medical technology (Petersen 2001, White 2001, Bubela and Caulfield 2004). Therefore, potential research participants may have skewed understandings of the benefits of genetic research, which would be reinforced by language that allows them to develop their own interpretations.

The language gap between researcher and potential research participants is obviously problematic for informed consent, but the solution is not straightforward. When jargon is the source of miscommunication, the problematic terms are relatively easy to identify and define. When the issue is vague but common language, almost every word could be considered for definition. There is a risk that one could over-reach, defining too many concepts, and rendering consent forms unbearably long and burdened with unreadable 'legalese'. Such a consent form is likely to be discarded rather than read, and is therefore of no value to the informed choice process.

On the other hand, some concepts must be defined if informed consent is going to truly inform. Presently, the best balance between ambiguity and precision is not clear. It seems

evident that concepts central to the research ought to be defined. In genetics, for example, it seems that “causation”, “gene”, “health”, “disease”, and “family” represent a few core concepts that would warrant definition. Unfortunately, however, there is no clear mechanism that identifies concepts requiring definition. Words that seem obvious to the authors may be understood differently by the audience. Some concepts requiring definition can probably be identified by the researchers, perhaps with the help of a research ethics board, and we should strive to ensure these definitions are provided. However, I believe the ideal approach will have to involve feedback from primary stakeholders, who could be involved in the process of developing informed consent documents and helping to identify language gaps in advance.

That varying understandings of common words could impact informed consent is a concept that I have not found discussed elsewhere. Thus, we should not expect that the authors of consent documents would take such variability into account. Although I have stated that such language may entice participation in genetics research, I do not wish to imply that researchers are manipulating their research participants. These researchers voluntarily participated in this ethics project. They are highly respected clinicians with their patients’ and research participants’ best interests at heart. They have labored to describe their research in simple and recognizable language. However, this analysis has demonstrated that because simple language can provoke numerous interpretations, miscommunication may still be possible. Therefore, potential research participants may be enticed into participation not through the actions of the researchers, but as a

consequence of differing beliefs and values being expressed using indistinguishable words.

Social impacts of genetics language and the language gap

The science of genetics has undoubtedly improved understanding of many conditions, leading to better communication among scientists, physicians, and patients, and potentially helping in decisions about prevention, screening, and treatment. However, considerable concerns have been raised about the impacts of genetics and society, many of which arise from questions of definition (Lippman 1991, Lippman 1992, Conrad and Gabe 1999, Shakespeare 1999, Wilson 2001, Knoppers and Chadwick 2005, Miller, Begbie et al. 2006, Mykitiuk and Nisker 2010). Ought condition X be defined by genetics, or is it better defined in terms of social determinants of health? Is condition Y really a "disease" or just part of the normal spectrum of health? Should an individual be defined and labeled just because they possess a specific mutation, gene, or physical attribute?

Through this analysis I recognized the rather evident fact that the authors of research documents have definitional control over key concepts, such as health, disease, and normalcy. In the studied documents, several basic concepts were noted to be redefined in unusual ways. "Illness" did not require signs or symptoms. Genetic relationships became a requisite for "families". Through the use of the phrase "disease gene", the authors fundamentally redefine disease as a genetic process. The definitional power of the author is represented in its extreme in the meshed concept "disease individual". This phrase

combines a single characteristic with the essence of an individual. The author claims the power to choose a potentially trivial feature, and make it an essential component of one's identity.

The consequences of definitional power are potentially profound. In the phrase "disease individual", the author characterizes an individual by a trait of the author's choosing, allowing a single trait to synechdochically represent the entire person (Parens and Asch 1999). This language could debase one's self worth and fundamentally change how one is viewed by others. It sets a tone of social acceptability (Rothstein 1991). Furthermore, it could alter personal and societal views on responsibility and blame for health conditions (Rothstein and Billings 1992, Crosseley 2002).

The redefinition of "family" based on genetic rather than social relationships could also have significant impacts. Individuals without specific mutations or diseases may have difficulty claiming membership in "disease families". The language used could alter individuals' understandings of their relationships by promoting the view that some bonds are more legitimate than others (Burgess 2001, Cox and Starzomski 2004, Holtzman 2008). This language also reinforces stigma surrounding childlessness, adoption, and infertility (Holtzman 2008). Finally, the genetic definitions promoted could have significant consequences for family law (Caulfield 2002, Holtzman 2008, Pioro 2008).

Similarly, defining disease as fundamentally genetic has significant social impacts. Other factors that influence health and disease, such as social and physical environments,

economics, personal behaviours, and available health services are ignored by this definition (Lippman 1991, Conrad and Gabe 1999, Mykitiuk and Nisker 2010). This language could influence decision making, allowing "genetic fatalism" to become a driving force in individual health decisions (Lippman 1992, Crosseley 2002). The use of such language by respected researchers will reinforce the widespread "genohype" we have already witnessed in the media (Petersen 2001, White 2001, Bubela and Caulfield 2004). The term "disease gene" implies that genetics offers the only, or the best, approach to studying and treating disease. Thus, this language will also influence decisions to fund, conduct, and participate in future research, potentially diverting money and resources away from important non-genetics enterprises (Lippman 1991, Wilson 2001, Mykitiuk and Nisker 2010).

Although I believe that the language used here illustrates the tremendous power that can be associated with language choice, I do not believe the authors are deliberately using language this way. In fact, given the subtle ways in which language's power is manifested, I imagine the authors were unaware of the potential social implications of their words. The definitional power that I believe these authors possess is not about choosing one advantageous meaning amongst many, but rather using a word without recognizing that it may have different meanings.

The existence of multiple understandings of concepts like health, disease, and normalcy is not intrinsically problematic. In fact, the multiplicity of understandings is probably beneficial when acknowledged and exploited as a source of creativity and balance in

medical research and health promotion. However, a concerning language gap can develop when certain definitions are allowed to dominate while other are ignored.

In these documents, the language of health was distinctly focused on genetics. Although a substantial body of literature finds fault with using genetics alone to define health and disease, these dissenting views are neither accessible to the audience through these documents, nor are they widely acknowledged or presented in the media. Therefore, potential research participants may be unaware that more than one definition of health and disease exists.

Even when multiple definitions are acknowledged, certain definitions may dominate, isolating minority views. Definitions develop through shared understandings within a group of language users. Broader understanding is the result of one group's definition winning out over others. In this way, definitions can become dominant because of the authority of a group of language users rather than any intrinsic value or sense in the language itself (Wittgenstein 1968, Holmes 2001). As a society, we value doctors and scientists as authority figures. As such, there is a risk that we may accept the genetic definitions used here as objectively and authoritatively true, instead of one of many possibilities.

That one view of health and disease is dominant in these documents is not problematic. The authors do not conceal other beliefs, they simply write with their own inherent and unrecognized biases – as we all do every time we use language. The potential problem

arises when the reader is not exposed to other views, or when the familiar language conceals the possibility that author and reader may disagree. The problem is not with the language these scientists used, but rather with a general unawareness that words can have different meanings for different groups of language users.

Language can be a powerful tool and accepted definition can have important emotional, social, and legal consequences. The definitions of words are neither static nor objective. We should anticipate that important social concepts, such as "disease" or "family", will be understood differently by different people. It is not problematic for groups to disagree about definitions; in fact, it could be beneficial. However, we ought to be wary of the power of one definition to eclipse all others and become the solely recognized, "true" definition.

The harms of language

The capacity of language to cause harm is closely related to the power language can have in social systems. Language forms a foundation of social acceptability, and inappropriate words can set the stage for harms. Words can scare, isolate, and dehumanize (Rothstein 1991, Hodgson, Hughes et al. 2005, Nisker and Daar 2006). Unfortunately, potentially but unintentionally harmful or pejorative language was seen throughout these documents.

Language can be offensive because of its inherent negativity and devaluation of characteristics that an individual does not have, but rather is (Shakespeare 1999, Byron, Cockshott et al. 2005). The inherent negativity of "risk" is an example that was

frequently seen. Other words, such as “disease”, “disorder”, and “malformation”, have likewise been classified as offensive and were frequently used here (Hodgson, Hughes et al. 2005). The word “mutation” is an example of language that is harmful because of its associated imagery (Condit, Archer et al. 2002, Condit, Dubriwny et al. 2004, Hodgson, Hughes et al. 2005). Finally, in allowing potentially trivial characteristics to stand for individuals, some observed language usages may debase a person’s intrinsic value (Parens and Asch 1999). This was seen with the interchangeable use of “disease” and “phenotype”, the discussion of “characterizing” individuals by their traits, and the use of disease names to refer to people. Similarly, but perhaps more concerning, is the phrase “disease individual” that fundamentally combines the individual with a disease, presumably diminishing any other aspects of his individuality.

The capacity of any individual word to cause harm was amplified by the interchangeable use of language. By simple probabilities, highly variable language increases the chance that pejorative language will be used. However, more troubling is the single common sense that all words adopt when used interchangeably. “Disease” is generally seen as a negative concept, and is therefore potentially pejorative in genetics, whereas “condition” is not (Hodgson, Hughes et al. 2005). However, in the studied documents, these terms were used interchangeably. The result is that both probably assume an offensive tone. Furthermore, the overall language may seem negative, as the author does not recognize any difference between these concepts.

One of the biggest challenges I faced in understanding the language of these documents was reconciling the language I observed with the stated goals of the documents and their authors. The language could be potentially harmful, but it was coming from a source that I am certain did not want to cause harm. These documents are written by doctors and medical researchers – individuals who spend their lives trying to help people. The studied documents were voluntarily contributed for analysis by scientists who, in participating, obviously value ethics and informed consent. The research grants were designed with the goal of improving health for the very people that the language appears to harm.

The solution to this apparent conflict, I believe, lies in the language gap. Words that appear innocuous to the author are potentially interpreted as harmful by the audience. The author uses these words without ever being aware of the other possible meanings. The harm is unintentional. I remember writing many grade school history papers blissfully unaware that my use of "Indian" or "Eskimo" was highly offensive. Similarly, I was shocked to see that the word "disease" was considered pejorative by some (Hodgson, Hughes et al. 2005). In medicine, I have used this word every day, and it was an immense struggle (that I have probably failed) to eliminate it from this document. Like me, I imagine that the respected scientists who authored these documents were never aware of this alternate understanding of "disease". Thus, the language gap does not just cause miscommunication, but it may also cause harm.

Potential Limitations

A potential limitation of this study is that only written language was analyzed. Written documents only represent a subset of the communication that is involved in the informed choice process. Potential research participants are provided with verbal information by the clinical research coordinator, and are provided with the opportunity to ask questions. Information letters and consent forms are provided as consistent and lasting sources of key information. Unlike clinical interactions, consent documents are reviewed many times, by their authors as well as research ethics boards, before being presented to potential research participants. Consequently, the language is more carefully selected than is possible in verbal conversation. Furthermore, informed consent documents are written with multiple goals, and heavily influenced by legal concerns, which may create unnatural or confusing language that would not be used in verbal conversation.

Written documents were primarily chosen because of their enduring nature, which allows them to be studied much more easily than clinical conversations. However, written documents are likely good representations of the general lexicon of an author and the general content of informed consent. The information provided during verbal consent is similar to that in written documents, although it is much more variable (Bjorn, Rossel et al. 1999, Sankar 2004). Thus, it is reasonable to assume that the word choice seen here would be mostly mirrored in clinical encounters. If anything, problems recognized in formal, polished written language are likely to be magnified in informal, uncensored clinical conversations.

Although my focus was the impacts of language on informed consent, the language that was analyzed and presented was drawn from a large variety of documents, including many that would not be available to potential research participants. Indeed, several language usages that were presented as potentially harmful or confusing were never found on information letters or consent forms. Thus, the relevance to informed choice of observations based on this language may be questioned.

I believe that the inclusion of research documents other than the information letter and consent form was essential to this analysis. The observation that language differed between consent documents, grant applications, and published papers increased the relevance of the language that was not seen by potential research participants. It illustrates the potential language gap between the researcher and potential research participant. It raises the question of why some words are used in one context while other words are used in another. Furthermore, the use of such language in any document illustrates that it is part of the author's lexicon, and therefore may be expected to arise in encounters with potential research participants.

Another important limitation of this study, which has been discussed throughout, is the lack of involvement of both researchers and potential research participants. The interpretations, therefore, represent only those of the analysts, and are not related to the intended meanings of the authors or the actual interpretations of potential research participants. However, the inductive research done here is an important first step that

identifies potentially troubling usages of language, which can later be more fully explored with researchers and potential research participants.

I must emphasize that the possible interpretations depicted here do not represent the intended meanings of the authors. I have focused on language with the potential to confuse or harm because of its importance for informed choice, but I do not believe that the authors intentionally used harmful or confusing language. These authors voluntarily sent me their documents for analysis. Furthermore, the consent forms would have been reviewed by local research ethics boards, and therefore others had previously been deemed the language appropriate. This apparent contradiction is explained by the language gap that can result from the multiple potential and unintended interpretations of language.

Finally, I must mention the general limitations of qualitative methodology. These are not truly limitations, but rather assumptions inherent to the qualitative paradigm. The results presented here are not generalizable to other research settings, but inductive descriptions of language use, rather than generalizability, was the goal of this research. Neither are the results objective, as the subjectivity of the researcher is an inherent component of the research (Lincoln and Guba 1985, Catanzaro 1988, Altheide 1996, Berg 2007). This study represents an important, initial exploration of informed choice language that inductively developed themes that will hopefully be examined further to determine their applicability in other research settings, their meaning in this particular research setting, and their true impact on informed consent.

Future directions

Language develops meaning only through the shared understandings of its users (Wittgenstein 1968). This study examined the language in documents, but never directly involved those who use the documents to communicate. Therefore, although I can describe language use, I cannot portray the authors' intended meanings or the interpretations of the audience. I have described being confused by many words with unclear meaning and variable use. I have highlighted terms that appear to assume new meanings in the context of genetics. I have indicated some language that could cause harm. However, it is not clear that this language actually confuses the intended audience, that it is the basis of any miscommunication, or that it causes any real harm. The nature and extent of a language gap cannot be established without the involvement of both the authors of the documents and their intended audience.

A true understanding of the language gap and the impacts of language on informed consent will require an appreciation of the use and understanding of language by the involved communities. In genetics, these communities include patients, their families, research participants, researchers, doctors and other health care providers, parents and potential parents, individuals living with genetic conditions or born using genetic technologies, and individuals hoping to use such technologies. Through shared experiences and connected social networks, each group will develop collective understandings of words like health, disease, and normalcy. Understandings will vary among groups, either on the basis of unnoticed differences in life experiences or as the result of more cognizant shared beliefs (Wittgenstein 1968, Holmes 2001, Stead, Eadie et

al. 2005). The in-depth interview based research that will be required to appreciate such differences in language use and understanding was an anticipated future step from the outset of this project.[27]

However, before interpretations of language can be explored, one must first determine the language of interest. Inductive qualitative research is essential when approaching a topic anew. It allows for novel insight and hypothesis development that is grounded in data. Previous research on language use in informed consent has focused on complicated scientific topics, complex language, or jargon as sources of miscommunication. Analysis of documents has therefore focused on classifying and quantifying previously identified problematic language (Waggoner and Sherman 1997, Davis, Holcombe et al. 1998, Bjorn, Rossel et al. 1999, Coyne, Xu et al. 2003, Stead, Eadie et al. 2005). On the other hand, the inductive methodology of qualitative content analysis provided the opportunity to recognize the language gap – a potentially significant obstacle to informed consent that I have not seen described elsewhere. Numerous specific language usages were identified as potential sources of a language gap, the meanings and understandings of which can now be explored in further detail with the various interested groups of language users.

I have focused on the gap in language that may occur when different groups interpret the same words differently. In the examples I have provided, it is possible that a true language gap does not exist. That is, despite the spectrum of possible interpretations, both

[27] How do researchers describe a "disease individual"? How do research participants react to this term? In-depth, structured interviews with all key populations of language users about each problematic language usage described here will be necessary to fully understand the intended meanings, understandings, and impacts of this language.

researcher and potential research participant could interpret all words identically. (Though, this seems unlikely given the numerous examples of concepts that are dramatically altered within the framework of genetics.) However, even if a true language gap does not exist in this case, the concept remains important, as the potential for a language gap always exists. Whenever we communicate, especially when communication occurs between groups with different backgrounds and life experiences, the potential for unrecognized miscommunication exists. Thus, when precision is important, such as in the context of informed consent to genetics research, the potential for a language gap must be considered.

Language is the principal tool available to ensure not just informed consent in research, but understanding in all health care interactions. The language of genetics is particularly complex, as it incorporates the unfamiliar concepts of prediction, probability, and heredity into health care discussions. However, the general difficulties with language described here are probably not unique to genetics. Many medical concepts are complex and will likely be presented with vague and variable language. Understandings of health related terms will vary, as scientists, physicians, and patients all have different experiences that shape their understandings. Therefore, we should anticipate that significant language gaps might be present in all medical communication.

Conclusion

This study suggests the potential for unrecognized miscommunication, or a language gap, in consent documents for genetic studies, but further work will be required to delineate

the scope, impact, and possible solutions to this problem. Meanwhile, the best approach to avoiding miscommunication resulting from a language gap is not clear. Until we understand more, I believe that the best approach is consistency in language use. By limiting the number of different words used to describe key concepts, the chance of divergent or harmful interpretations decreases. Furthermore, consistent language may act as a signal of a concept's significance. Rather than being overwhelmed and confused by complex and variable language, potential research participants will be more likely to recognize key terms, even if they are being used in novel or unfamiliar ways. This provides the opportunity to seek clarification. Additionally, the author of the consent document should identify key concepts at the outset, in order to ensure consistency, and therefore will be more likely to provide explicit definitions or opportunities for explanation.

However, consistency does not eliminate the language gap. The meanings of common words often seem obvious. The consistent use of a word like "disease" throughout a consent form does not guarantee that the researcher and potential research participants agree on its meaning. However, the word is so familiar that its definition is unlikely to be discussed. This problem can be partially rectified by identifying and defining key terms in advance. Unfortunately, it is not prospectively obvious which words might generate a language gap, and therefore require definition. Assessment of differences in language interpretation requires the involvement of individuals from each different group of language users involved. Therefore, I believe that a feedback process in document development will prove important in diminishing the impacts of the language gap. One

solution would be to involve primary stakeholders from the proposed community of research participants in the writing of research documents. This would allow for the discussion of key terms and identification of possible sources of miscommunication before informed consent documents are used with potential research participants.

There is a lot to learn about the impacts of the language gap on communication. As we continue to improve medical communication by favouring plain language over jargon, the language gap may become a major source of miscommunication. Plain language is an excellent goal, but the language gap reminds us that it may not be enough. We must ensure that simplified language is still able to describe the complexities of science. And we must recognize that for plain language to truly improve comprehension, individuals must understand not just the words, but the underlying concepts.

GLOSSARY

A glossary was a late addition to this thesis. Although I wrote about the problems that can result from a lack of clear definitions, the task of providing definitions was daunting. I did not know how to begin – a glossary is not a standard component of academic writing and I had never been taught how to develop one. It wasn't clear exactly what terms required definitions, how they ought to be defined, and by whom.

I believe that key concepts must be defined, and that is what I will attempt to accomplish below. However, it is important to recognize that this list was developed entirely by me. I cannot be certain which terms may be understood differently than my intended meaning. Ideally, the audience would be somehow involved in determining what concepts should be defined. I believe this highlights another responsibility of the author: the need for consistency. It is only through the consistent use of language that the audience will be able to recognize the terms that they would like better defined.

* I have provided definitions of a number of genetics related terms. This is obviously problematic in a project that was designed to examine the understandings and meanings of these words in genetics research documents. I provide my definitions not as 'the correct definition', nor as the definitions used within these documents. The definitions provided here are meant to provide the reader with a sense of my understandings of genetics, and therefore my potential biases.

Allele*: Any subset of DNA order within a gene. Alleles are what actually represent the differences between us at the genetic level. We all share the same genes, but within each gene, there are many possible DNA sequences, which are called alleles.

Anonymization: The process of removing information that could identify research participants from presented data so that research participants remain anonymous.

Community of language users: Any group of individuals with meaningful interactions or shared life experiences that therefore develop a shared use of language.

Conceptual meshing: A linguistic formulation observed frequently in the studied documents, in which two distinct concepts are combined into a single new entity, often without clear meaning.

Definition: A method of formalizing a word's intended meaning. When a word is defined, each and every time that word is used it is intended to mean the same thing.

Disease individual: A novel phrase seen here that combines disease and individual into a single new concept without clear meaning.

Explanation: Any statement that attempts to put a complex concept into plain words.

Family*: A group of closely related individuals. I have not defined the nature of the relationship that is required to count as family. For now, I would say that any group of individuals that self-identifies as a family is a family.

Gene*: A stretch of DNA (deoxyribonucleic acid) responsible for encoding a specific RNA (ribonucleic acid) molecule.

Genome*: The entire collection of a single individual's or organism's genetic material.

Identifier: Any piece of data that could, either alone or when combined with other data, expose the identity of a research participant.

Language: Any system used to communicate. Here, I have exclusively discussed word based language, specifically English. However, language could also involve other signals or gestures, such as American Sign Language.

Language gap: The divide that exists between two individuals when they use the same language (the same words or signals) with different meanings and understandings.

Meaning: The message or image that is intended to be conveyed by when using a word or symbol is used.

Mutation*: Any change in the sequence of an organism's DNA (deoxyribonucleic acid).

Phenotype*: The observable characteristics of an organism.

Understanding: The message or image that is interpreted when a word or symbol is encountered.

Vague language: Language that can be interpreted in many different ways. Note: This term was often used interchangeably with "ambiguous language" in these documents.[28]

[28] I have argued for consistency in language use throughout this paper. Ideally, I would have liked to have used a single term to describe this concept. Unfortunately, I recognized this dual usage only after finishing the thesis, and modification was impractical.

The X gene or the disease gene: A frequently used phrase that combines a specific disease name, such as cancer or diabetes, with the word gene, that lacks a clear scientific meaning. Examples include ‘the obesity gene’, ‘the breast cancer gene’, and ‘the cystic fibrosis gene’.

REFERENCES

Alleyne, R. (2010). 'Drunken' gene discovered by scientists. The Telegraph. London.

Alter, B. P., P. S. Rosenberg and L. C. Brody (2007). "Clinical and molecular features associated with biallelic mutations in FANCD1/BRCA2." Journal of Medical Genetics **44**: 1-9.

Altheide, D. L. (1996). Qualitative Media Analysis. Thousand Oaks, Sage Publications.

The American Heritage Dictionary. (2004). Condition. The American Heritage® Dictionary of the English Language, Fourth Edition, Houghton Mifflin Company.

Annas, G. J. (2001). "Reforming informed consent to genetic research." Journal of the American Medical Association **286**(18): 2326-2328.

Appelbaum, P. S. (2002). "Clarifying the ethics of clinical research: A path toward avoiding the therapeutic misconception." American Journal of Bioethics **2**(2): 22-25.

Appelbaum, P. S., L. H. Roth, C. W. Lidz, P. Benson and W. Winslade (1987). "False hopes and best data: Consent to research and the therapeutic misconception." The Hastings Center Report **17**(2): 20-24.

Baevsky, R. (2008). "Speaking in plain English." Annals of Emergency Medicine **54**(4): 450-451.

Baker, M. T. and H. A. Taub (1983). "Readability of informed consent forms for research in a Veterans Administration medical center." Journal of the American Medical Association **250**: 2646.

Balmana, J., O. Diez and M. Castiglione (2009). "BRCA in breast cancer: ESMO clinical recommendations." Annals of Oncology **40**(S4): 19-20.

Bayertz, K. (1998). "What's special about molecular genetic diagnostics?" Journal of Medicine and Philosophy **23**(3): 247-254.

Baylis, F., Downie, J., Hoffmaster, B., Sherwin, S., Ed. (2004). Health care ethics in Canada. Toronto, Thomson Nelson.

Beardsley, E., M. Jefford and L. Mileshkin (2007). "Longer consent forms for clinical trials compromise patient understanding: So why are they lengthening?" Journal of Clinical Oncology **25**(9): E13-E14.

Beauchamp, T. and J. Childress (2001). Principles of Biomedical Ethics. Oxford, Oxford University Press.

Beaulieu, A. and A. Lippman (1995). ""Everything you need to know": how women's magazines structure prenatal diagnosis for women over 35." Women Health **23**(3): 59-74.

Bedell, S. E., T. B. Graboys, E. Bedell and B. Lown (2004). "Words that harm, words that heal." Archives of Internal Medicine **164**: 1365-1367.

Benkendorf, J., M. Prince, M. Rose, A. De Fina and H. Hamilton (2001). "Does indirect speech promote nondirective genetic counseling? Results of a sociolinguistic investigation." American Journal of Medical Genetics **106**(3): 109-207.

Berg, B. L. (2007). Qualitative Research Methods for the Social Sciences. Boston, Allyn and Bacon.

Bhutta, Z. A. (2004). "Beyond informed consent." Bulletin of the World Health Organization **82**(10): 771-777.

Bjorn, E., P. Rossel and S. Holm (1999). "Can the written information to research subjects be improved? An empirical study." Journal of Medical Ethics **25**(3): 263-267.

Bubela, T. M. and T. A. Caulfield (2004). "Do the print media "hype" genetic research? A comparison of newspaper stories and peer-reviewed research papers." CMAJ **170**(9): 1399-1407.

Burgess, M. (2001). "Beyond consent: ethical and social issues in genetic testing." Nature Reviews: Genetics **2**(2): 147-151.

Burgess, M. (2007). "Proposing modesty for informed consent." Social Science and Medicine **65**: 2284-2295.

Burgess, M. M., C. M. Laberge and B. M. Knoppers (1998). "Bioethics for clinicians: 14. Ethics and genetics in medicine." Canadian Medical Association Journal **158**(10): 1309-1313.

Byron, M., Z. Cockshott, H. Brownett and T. Ramkalawan (2005). "What does 'disability' mean for medical students? An exploration of the words medical students associate with the term 'disability.'." Medical Education **39**: 176-183.

Catanzaro, M. (1988). Using Qualitative analytical techniques. Nursing Research: Theory and Practise. N. F. Woods, Catanzaro, M. Thousand Oaks, Sage Publications.

Caulfield, T. (2002). "A shifting concept of family?" Nature Reviews: Genetics **3**: 823.

Chapple, A., P. Campion and C. May (1997). "Clinical terminology: anxiety and confusion amongst families undergoing genetic counseling." Patient Education and Counseling **32**: 81-91.

Chenail, R. J. (1995). "Presenting qualitative data." The Qualitative Report **2**(3).

Churchill, L. R., M. L. Collins, N. M. P. King, S. G. Pemberton and K. A. Wailoo (1998). "Genetic research as therapy: Implications of "gene therapy" for informed consent." Journal of Law, Medicine, and Ethics **26**(1): 38-47.

Condit, C. M., P. J. Archer and E. Sefcovic (2002). "The changing meanings of "mutation:" A contextualized study of public discourse." Human Mutation **19**: 69-75.

Condit, C. M., T. Dubriwny, J. Lynch and R. Parrott (2004). "Lay people's understanding of and preference against the word "mutation"." American Journal of Medical Genetics **130a**: 245-250.

Connor, S. (1995). Research confirms 'gay gene' discovery. The Independent. London.

Conrad, P. (1997). "Public eyes and private genes: Historical frames, new constructions, and social problems." Social Problems **44**(2): 139-154.

Conrad, P. (1999). "A mirage of genes." Sociology of Health & Illness **21**(2): 228-241.

Conrad, P. and J. Gabe (1999). "Sociological perspectives on the new genetics: an overview." Sociology of Health and Illness **21**(5): 505-516.

Conrad, P. and D. Weinberg (1996). "Has the gene for alcoholism been discovered three times since 1980? A news media analysis." Perspectives on Social Problems **8**: 3-24.

Corrigan, O. (2003). "Empty ethics: The problem with informed consent." The Sociology of Health and Illness **25**(3): 768-792.

Corti, L., A. Day and G. Backhouse (2000). "Confidentiality and informed consent: Issues for consideration in the preservation of and provision of access to qualitative data archives." Forum: Qualitative Social Research **1**(3).

Cox, G., J. Burger, V. Lip, U. S. Mau, K, B. Wu and B. Horsthemke (2002). "Intracytoplasmic sperm injection may increase the risk of imprinting defects." American Journal of Human Genetics **71**(1): 162-164.

Cox, S. M. and R. C. Starzomski (2004). "Genes and geneticization? The social construction of autosomal dominant polycystic kidney disease." New Genetics and Society **23**(2): 137-166.

Coyne, C., R. Xu, P. Raich, K. Plomer, M. Dignan, L. Wenzel, D. Fairclough, T. Habermann, L. Schnell, S. Quella and D. Cella (2003). "Randomized, controlled trial of an easy-to-read informed consent statement for clinical trial participation: A study of the Eastern Cooperative Oncology Group." Journal of Clinical Oncology **21**(5): 836-842.

Crosseley, M. L. (2002). "'Could you please pass one of those health leaflets along?': exploring health, morality and resistance through focus groups." Social Science and Medicine **55**: 1471-1483.

Crusio, W. E. (2002). "My mouse has no phenotype." Genes, Brain and Behavior **1**(2): 71.

Davis, T., R. Holcombe, B. HJ, S. Pramanik and S. Divers (1998). "Informed consent for clinical trials: a comparative study of standard versus simplified forms." Journal of the National Cancer Institute **90**(9): 668-674.

de Crespigny, L. (2003). "Words matter: Nomenclature and communication in perinatal medicine." Clinics in Perinatology **30**: 17-25.

Deech, R. (1998). "Family law and genetics." Modern Law Review **61**(5): 697.

Devlieger, P. J. (1999). "From handicap to disability: language use and cultural meaning in the United States." Disability and Rehabilitation **21**(7): 346-354.

Devlin, K. (2008). Six new obesity genes discovered. The Telegraph. London.

Duster, T. (1990). Backdoor to Eugenics. New York, Routledge.

Erby, L. H., D. Roter, S. Larson and J. Cho (2008). "The rapid estimate of adult literacy in genetics (REAL-G): A means to assess literacy deficits in the context of genetics." American Journal of Medical Genetics Part A **146A**(2): 174-181.

Faden, R. R. and T. Beauchamp (1986). A history and theory of informed consent. New York, Oxford Univsersity Press.

Farrell, M., L. Deuster, J. Donovan and S. Christopher (2008). "Pediatric Residents' Use of Jargon During Counseling About Newborn Genetic Screening Results." Pediatrics **122**(2): 243-249.

Fleischman, S. (1999). "I am..., I have..., I suffer from...: A linguist reflects on the language of illness and disease." Journal of Medical Humanities **20**(1): 3-32.

Flory, J. and E. Emanueal (2004). "Interventions to improve research participants' understanding in informed consent for research: A systematic review." Journal of the American Medical Association **292**(13): 1593-1601.

Gallo, C., F. Perrone, S. De Placido and C. Giusti (1995). "Informed versus randomised consent to clinical trials." Lancet **346**(8982): 1060-1064.

Gannett, L. (1999). "What's in a cause? The pragmatic dimensions of genetic explanations." Biology and Philosophy **14**: 349-374.

Ghulam, A. T. (2006). "Patients' satisfaction with the preoperative informed consent: A multi-centre questionaire survey in switzerland." Mayo Clinic Proceedings **81**(3): 307-312.

Gigerenzer, G., W. Gaissmaier, E. Kurz-Milcke, L. M. Schwartz and S. Woloshin (2008). "Helping doctors and patients make sense of health statistics." Psychological science in the public interest **8**(2): 53-96.

Giusti, R. M., Rutter, J. L., Duray, P. H., Freedman, L. S., Konichezky, M., Fisher-Fischbein, J., Greene, M. H., Maslansky, B., Fischbein, A., Gruber, S. B., Rennert, G., Ronchetti, R. D., Hewitt, S. M., Struewing, J. P., Iscovich, J. (2003). "A twofold increase in BRCA mutation related prostate cancer among Ashkenazi Israelis is not associated with distinctive histopathology." Journal of Medical Genetics **40**: 787-792.

Goldstein, A. O., P. Frasier, P. Curtis, A. Reid and N. E. Kreher (1996). "Consent for readability in university-sponsored research." Journal of Family Practice **42**(6): 606-611.

Gower, T. (2000). "Are Some Guys Just Born Reckless?" The Los Angeles Times.

Graneheim, U. H. and B. Lundman (2004). "Qualitative content analysis in nursing research: concepts, procedures, and measures to achieve trustworthiness." Nurse Education Today **24**: 105-112.

Groce, N. (1985). Everyone Here Spoke Sign Language: Hereditary Deafness on Marthas Vineyard. Cambridge, Harvard University Press.

Grunder, T. M. (1980). "On the readability of surgical consent forms." The New England Journal of Medicine **302**: 900-902.

Guyatt, G., V. Montori, P. J. Devereaux, H. Schunemann and M. Bhandari (2004). "Patients at the centre: In our practice, and in our use of language." Evidence-Based Medicine **9**(1): 6-7.

Hall, S. S. (2006). "Exploring young adults' belief systems about marriage." Journal of Family Issues **27**(10): 1437-1458.

Hallowell, N. (1999). "Doing the right thing: genetic risk and responsibility." Sociology of Health & Illness **21**(5): 597-621.

Hallowell, N. and M. Richards (1997). "Understanding life's lottery: an evaluation of studies of genetic risk awareness." Journal of Health Psychology **2**(1): 31-43.

Hammerschmidt, D. E. and M. A. Keane (1992). "Institutional review board (IRB) review lacks impact on readability of consent forms for research." American Journal of the Medical Sciences **304**(6): 348-351.

Harmon, A. (2007). Facing life with a lethal gene. New York, The New York Times.

Harris, J. (2000). "Is there a coherent social conception of disability?" Journal of Medical Ethics **26**: 95-100.

Henderson, G. E., A. M. Davis, N. M. P. King, M. M. Easter, C. R. Zimmer, B. B. Rothschild, B. S. Wilfond, D. K. Nelson and L. R. Churchill (2004). "Uncertain benefit: Investigators' views and communications in early phase gene transfer trials." Molecular Therapy **10**: 225-231.

Heritage, J., J. D. Robison, M. N. Elliott, M. Beckett and M. Wilkes (2007). "Reducing patients' unmet concerns in primary care: the difference one word can make." Journal of General Internal Medicine **22**(10): 1429-1433.

Hochhauser, M. (1997). "Some overlooked aspects of consent form readability." IRB: Ethics and Human Research **19**(5): 5-9.

Hochhauser, M. (2000). "The informed consent form: Document development and evaluation." Drug Information Journal **34**(4): 1309-1317.

Hodgson, J., E. Hughes and C. Lambert (2005). ""SLANG" - Sensitive Language and the New Genetics - An exploratory study." Journal of Genetic Counseling **14**(6): 415-421.

Hoedemaekers, R., B. Gordijn, Y. Hekster and F. van Agt (2006). "The complexities of ethical evaluation of genomics research." HEC Forum **18**(1): 18-36.

Hoeyer, K. and N. Lynoe (2006). "Motivating donors to genetic research? Anthropological reasons to rethink the role of informed consent." Medicine, Health Care and Philosophy **9**: 13-23.

Holmes, J. (2001). An introduction to sociolinguistics. Toronto, Pearson Education Ltd.

Holtzman, M. (2008). "Defining family: Young adults' perceptions of the parent-child bond." Journal of Family Communication **8**(3): 167-185.

Homan, R. (1991). The ethics of social research. New York, Longman.

Hopper, K. D., T. R. TenHave and J. Hartzel (1995). "Informed consent forms for clinical and research imaging procedures: How much do patients understand?" American Journal of Roentgenology **164**(2): 493-496.

Hsieh, H. F. and S. E. Shannon (2005). "Three approaches to qualitative content analysis." Qualitative Health Research **15**(9): 1277-1288.

Hull, S. C., K. Glanz, A. Steffen and B. S. Wilfond (2004). "Recruitment approaches for family studies: attitudes of index patients and their relatives." IRB: Ethics and Human Research **26**(4): 12-18.

The Human Genome Project Completion: Frequently asked questions. (2008). "The Human Genome Project Completion: Frequently asked questions." Retrieved November 16, 2008, from http://www.genome.gov/11006943.

Issa, A. M. (2002). "Ethical perspectives on pharmacogenomic profiling in the drug development process." Nature Reviews: Drug Discoveries **1**: 300-308.

Jefford, M. and R. Moore (2008). "Improvement of informed consent and the quality of consent documents." Lancet Oncology **9**: 485-493.

Jette, A. J. (2005). "The changing language of disablement." Physical Therapy **85**(2): 118-119.

Jette, A. J. (2006). "Toward a common language for function, disability, and health." Physical Therapy **85**(5): 726-734.

Kay, L. E. (1995). "Who wrote the book of life? Information and the transformation of molecular biology, 1945-55." Science in Context **8**(4): 609-634.

Keller, E. (2000). The Century of the Gene. Cambridge, Harvard University Press.

Kent, G. (1996). "Shared understandings for informed consent: The relevance of psychological research on the provision of information." Social Science and Medicine **43**(10): 1517-1523.

Kerr, A., S. Cunningham-Burley and A. Amos (1998). "Eugenics and the new genetics in Britain: examining contemporary professionals' accounts." Science, Technology, and Human Values **23**(2): 175-198.

Kessler, S. (1990). "Current psychological issues in genetic counseling." Journal of Psychosomatic Obstetric Gynecology **11**(S): 5-18.

Kevles, D. and L. Hood, Eds. (1992). The Code of Codes: Scientific and Social Issues in the Human Genome Project Cambridge, Harvard University Press.

Kimmelman, J. and A. Levenstadt (2005). "Elements of style: Consent form language and the therapeutic misconception in phase 1 gene transfer trials." Human Gene Therapy **16**(4): 502-508.

King, N. M. P. (1999). "Rewriting the "points to consider": The ethical impact of guidance document language." Human Gene Therapy **10**(1): 133-139.

King, N. M. P., G. E. Henderson, L. R. Churchill, A. M. Davis, S. C. Hull, D. K. Nelson, P. C. Parham-Vetter, B. Bluestone, B. B. Rothschild, M. M. Easter and B. S. Wilfond (2005). "Consent forms and the therapeutic misconception: The example of the gene transfer research." IRB: Ethics and Human Research **21**(1): 1-9.

Knoppers, B. M. (2005). "Consent revisited: Points to consider." Health Law Review **13**: 33-38.

Knoppers, B. M. and R. Chadwick (2005). "Human Genetic research: emerging trends in ethics." Nature Reviews: Genetics **6**(1): 75-79.

Kodish, E. D., M. Eder, R. B. Noll, K. Ruccione, B. Lange, A. Angiolillo, R. Pentz, S. J. Zyzanski, L. A. Siminoff and D. Drotar (2004). "Communication of randomization in childhood leukemia trials." Journal of the American Medical Association **291**(4): 470-475.

Kong, A., Barnett, G.O., Mosteller, F., Youtz, C. (1986). "How medical professionals evaluate expressions of probability." New England Journal of Medicine **315**: 741-744.

Levine, R. (1988). Ethics and Regulation of Clinical Research. New Haven, CT, Yale University Press.

Lincoln, Y. S. and E. G. Guba (1985). Naturalistic Inquiry. Thousand Oaks, Sage Publications.

Linell, P., V. Adelswrd, L. Sachs, M. Bredmar and U. Lindstedt (2002). "Expert talk in medical contexts." Research on Language and Social Interaction **32**(5): 195-218.

Lippman-Hand, A. and C. Fraser (1979). "Genetic Counseling: The provision and reception of information." American Journal of Medical Genetics **3**(2): 113-127.

Lippman, A. (1991). "Prenatal genetic testing and screening: Constructing needs and reinforcing inequities." American Journal of Law and Medicine **17**(1-2): 15-50.

Lippman, A. (1992). "Led (astray) by genetic maps: the cartography of the human genome and health care." Social Science and Medicine **35**(12): 1469-1476.

Liss, P., O. Aspevali, D. Karlsson and U. Forsum (2004). "Interpreting definitions: The problem of interpreting definitions of medical concepts." Medicine, Health Care and Philosophy **7**: 137-141.

Loftus, E. F. and J. C. Palmer (1974). "Reconstruction of automobile destruction: An example of the interaction between langauge and memory." Journal of verbal learning and verbal behavior **13**(5): 585-589.

LoVerde, M. E., A. V. Prochazka and R. L. Byyny (1989). "Research consent forms: Continued unreadability and increasing length." Journal of General Internal Medicine **4**: 410-412.

Lynoe, N., M. Sandlund, G. Dahlqvist and L. Jacobsson (1991). "Informed consent: study of quality of information given to participants in a clinical trial." British Medical Journal **303**: 610-613.

Makoul, G. (2001). "Essential elements of communication in medical encounters: the Kalamazoo consensus statement." Academic Medicine **76**(4): 390-393.

Manis, J. P. (2007). "Knock out, knock in, knock down - genetically manipulated mice and the Nobel prize." New England Journal of Medicine **357**(24): 2426-2429.

Marx, G. T. (1999). "What's in a name? Some reflections on the sociology of anonymity." The Information Society **15**: 99-112.

Miles, M. B. and A. M. Huberman (1994). Qualitative Data Analysis: An expanded sourcebook. Thousand Oaks, Sage Publications.

Miller, F., M. Begbie, M. Giacomini, C. Ahern and E. Harvey (2006). "Redefining disease? The nosologic implicaions of molecular genetic knowledge." Perspectives in Biology and Medicine **49**(1): 99-114.

Miller, F. A. and K. Alvarado (2005). "Incorporating documents in qualitative nursing research." Journal of Nursing Scholarship **37**(4): 348-353.

Murgatroyd, R. J. and R. M. Cooper (1991). "Readability of informed consent forms." American Journal of Hospital Pharmacy **48**(12): 2651-2652.

Murphy, K. M., Brune, K. A., Griffin, C., Sollenberger, J. E., Petersen, G. M., Bansal, R., Hruban, R. H., Kern, S. E. (2002). "Evaluation of candidate genes MAP2K4, MADH4, ACVR1B, and BRCA2 in familial pancreatic cancer: deleterious BRCA2 mutations in 17%." Cancer Research **62**: 3789-3792.

Mykitiuk, R. and J. Nisker (2010). The social determinant of "health" of embryos. The "Healthy" Embryo: Social, Biomedical, Legal, and Philosophical Perspectives. J. Nisker, F. Baylis, I. Karpin, C. McLeod and R. Mykitiuk (eds.). New York, Cambridge University Press.

Nathanson, K. and B. Weber (2001). ""Other" breast cancer susceptibility genes: searching for more holy grail." Human Molecular Genetics **10**(7): 715-720.

National Center for Biotechnology Information. (2008). The NCBI handbook [Internet]. National Center for Biotechnology Information. Bethesda, MD, National Library of Medicine (US).

Nelkin, D. and S. Lindee (2004). The DNA Mystique : The Gene as a Cultural Icon. Ann Arbor, University of Michigan.

Nespor, J. (2000). "Anonymity and plac in qualitative inquiry." Qualitative Inquiry **6**(4): 546-569.

Nisker, J. and A. S. Daar (2006). "Moral presentation of genetics-based narratives for public understanding of genetic science and its implications." Public Understanding of Science **15**: 113-123.

Nolan, M. T. (1999). "Consent documents, reproductive issues, and the inclusion of women in clinical trials." Academic Medicine **74**(3): 275-281.

O'Neill, O. (2003). "Some limits of informed consent." Journal of Medical Ethics **29**: 4-7.

Parens, E. and A. Asch (1999). "The disability critique of prenatal testing: Reflections and recommendations." Hastings Center Report **29**(5): S1-S22.

Parsons, E., Atkinson, P. (1992). "Lay constructions of genetic risk." Sociology of Health & Illness **14**: 437-455.

Passarge, E. (2007). Color atlas of genetics. New York, Stuttgart.

Paul, D. B. (1992). "Eugenic anxieties, social realities, and political choices." Social Research **59**(3): 21-31.

Petersen, A. (2001). "Biofantasies: Genetics and medicine in the print news media." Social Science and Medicine **52**: 1255-1268.

Pinker, S. (2005). Sniffing Out the Gay Gene. The New York Times. New York.

Pioro, M., Mykitiuk, R., Nisker, J (2008). "Wrongful birth litigation and prenatal screening." Canadian Medical Association Journal **179**(10): 1027-1030.

Pothier, D. D. (2005). "Many patients may not understand consent forms." British Medical Journal **330**: 1151.

Reid, S., Renwick, A., Seal, S., Baskcomb, L., Barfoot, R., Jayatilake, H., The Breast Cancer Susceptibility Collaboration (UK), Pritchard-Jones, K., Stratton, M. R., Ridolfi-Luthy, A., Rahman, N. (2005). "Biallelic BRCA2 mutations are associated with multiple malignancies in childhood including familial Wilms tumour." Journal of Medical Genetics **42**: 147-151.

Richards, M. (1993). "The new genetics: some issues for social scientists." Sociology of Health & Illness **15**(5): 567-586.

Rock, F. (2001). "Policy and practise in the anonymization of linguistic data." International Journal of Corpus Linguistics **6**(1): 1-26.

Roget's 21st Century Thesaurus. (2008). "Mutant." Roget's 21st Century Thesaurus, Third Edition Retrieved November 9, 2008, from http://dictionary.reference.com/browse/mutant.

Rothman, B. K. (1998). Genetic maps and human imaginations: The limits of science in understanding who we are. New York, W.W. Norton & Co.

Rothstein, J. M. (1991). "Sticks and stones." Physical Therapy **71**(7): 498.

Rothstein, M. A. and P. Billings (1992). "But is he genetically diseased?" Hastings Center Report **22**(4): S18-21.

Sandelowski, M. (1994). "Focus on qualitative methods: The use of quotes in qualitative research." Research in Nursing & Health **17**(6): 179-182.

Sankar, P. (2004). "Communication and miscommunication in informed consent to research." Medical Anthropology Quarterly **18**(4): 429-446.

Saraiva, J. M., E. Anionwu, M. Belo, T. Jenkins, U. Kristoffersson, I. Marques, H. G. Santos, J. Sequeiros, S. A. Simpson, D. Wertz, C. Monteiro, J. M. Saraiva, E. Anionwu, M. Belo, T. Jenkins, U. Kristoffersson, I. Marques, H. G. Santos, J. Sequeiros, S. A. Simpson, D. Wertz and C. Monteiro (2001). "Issues in human GenEthics." Genetics in Medicine **3**(3): 218-219.

Schuntermann, M. F. (2005). "The implementation of the international classification of functioning, disability and health in Germany: experiences and problems." International Journal of Rehabilitation Research **28**(2): 93-102.

Secko, D. M., N. Preto, S. Niemeyer and M. M. Burgess (2009). "Informed consent in biobank research: A deliberative approach to the debate." Social Science and Medicine **68**: 781-789.

Shakespeare, T. (1999). "'Losing the plot'? Medical and activist discourses on contemporary genetics and disability." Sociology of Health & Illness **21**(5): 669-688.

Slevin, M., J. Mossman, A. Bowling, R. Leonard, W. Steward, P. Harper, M. McIllmurray and N. Thatcher (1995). "Volunteers or victims: patients' views of randomised cancer clinical trials." British Journal of Cancer **71**(6): 1270-1274.

Speed, E. (2006). "Patients, consumers, and survivors: A case study of mental health service user discourse." Social Science and Medicine **62**: 28-38.

Stableford, S. and W. Mettger (2007). "Plain language: a strategic response to the health literacy challenge." Journal of Public Health Policy **28**: 71-93.

Stead, M., D. Eadie, D. Gordon and K. Angus (2005). ""Hello, hello - It's English I speak!": A qualitative exploration of patients' understanding of the science of clinical trials." Journal of Medical Ethics **31**(11): 664-669.

Stockdale, A. (1999). "Waiting for the cure: Mapping the social relations of human gene therapy research." Sociology of Health & Illness **21**(5): 579-596.

Strauss, A. and J. Corbin (1998). Basics of Qualitative Research: Techniques and procedures for developing grounded theory. London, Sage Publications.

Sugarman, J., N. E. Kass, S. N. Goodman, P. Perentesis, P. Fernandes and R. R. Faden (1998). "What patients say about medical research." IRB: Ethics and Human Research **20**(4): 1-7.

Sutherland, H., G. Lockwood, D. Tritchler, F. Sem, L. Brooks and J. Till (1991). "Communicating probabilistic information to cancer patients: is there 'noise' on the line?" Social Science and Medicine **32**(6): 725-731.

Sweeney, K. G., D. MacAuley and D. P. Gray (1998). "Personal significance: the third dimension." Lancet **351**(9096): 134-136.

Tarnoawski, K., D. Allen, C. Mayhall and P. Kelly (1990). "Readability of pediatric biomedical informed consent forms." Pediatrics **85**: 58-62.

Taylor, K. and R. Mykitiuk (2001). "Genetics, Normalcy, and Disability." ISUMA: Canadian Journal of Policy Research/Revue **2**(3): 65.

ten Have, H. (2003). "Genetic advances require comprehensive bioethical debate." Croatian Medical Journal **44**(5): 533-537.

Thomson, D., L. Bzdel, K. Golden-Biddle, T. Reay and C. A. Estabrooks (2005). "Central questions of anonymization: A case study of secondary use of qualitative data." Forum: Qualitative Social Research **6**(1).

Torres, J. M. (2002). "The importance of genetic services for the theory of health: A basis for an integrating view of health." Medicine, Health Care and Philosophy **5**: 43-51.

von den Hoonaard, W. C. (2003). "Is anonymity an artifact in ethnographic research?" Journal of Academic Ethics **1**: 141-151.

Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans. (2009). Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans, Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, Social Sciences and Humanities Research Council of Canada.

Waggoner, W. C. and D. M. Mayo (1995). "Who understands? A survey of 25 words of phrases commonly used in proposed clinical research consent forms." IRB: Ethics and Human Research **17**(1): 6-9.

Waggoner, W. C. and B. B. Sherman (1997). "Who understands? A survey of 27 words, phrases, or symbols used in proposed clinical research forms." IRB: Ethics and Human Research **18**(3): 8-10.

Wagley, P. (1987). "The Hippocratic Oath." Humane Medicine **3**(2): 110-114.

Walter, F. M., J. D. Emery, M. Rogers and N. Britten (2004). "Women's views of optimal risk communication and decision making in general practice consultations about the menopause and hormone replacement therapy." Patient Education and Counseling **53**: 121-128.

Wasserman, D., J. Bickenbach, R. Wachbroit and D. MacLean (2005). Quality of Life and Human Difference: Genetic Testing, Health Care, and Disability. Cambridge, Cambridge University Press.

Wertz, D. C. (2002). "Genetic discrimination - an overblown fear?" Nature Reviews: Genetics **3**(7): 496.

Wertz, D. C. (2003). "Ethical, social and legal issues in pharmacogenomics." The Pharmacogenetics Journal **3**: 194-196.

Wertz, D. C. and B. M. Knoppers (2002). "Serious genetic disorders: Can or should they be defined?" American Journal of Medical Genetics **108**: 29-35.

White, L., J. Jones, C. Felton and L. Pool (1996). "Informed consent for medical research: common discrepancies and readability." Accademic Emergency Medicine **3**(8): 745-750.

White, T. (2001). 'Get out of my lab, Lois!': in search of the media gene. Altered Genes II: the future? R. Hindmarsh, Lawrence, G. Melbourne, Scribe Publications.

Wilkinson, R. and M. Marmot (2003). Social determinants of health: The solid facts. Copenhagen, World Health Organization.

Williams, C. J. and M. Zwitter (1994). "Informed consent in European multicentre randomized control clinical trials - are patients really informed?" European Journal of Cancer **30**(7): 907-910.

Wilson, J. C. (2001). "Disability and the genome: resisting the standardized genomic text." Disability Studies Quarterly **21**(3): 166-179.

Wittgenstein, L. (1968). Philosophical investigations. Oxford, Basil Blackwell.

World Health Organization (1948). Constitution of the World Health Organization. Geneva.

Zhang, J. (1997). "The nature of external representations in problem solving." Cognitive Science **21**: 179-217.

Zhang, J. (2002). "Representations of health concepts: a cognitive perspective." Journal of Biomedical Informatics **35**: 17-24.

Zhang, J. and D. A. Norman (1995). "A representational analysis of numeration systems." Cognition **57**: 271-295.

CURRICULUM VITAE

Name:	Dr. Justin Morgenstern
Post-secondary Education:	Queen's University Kingston, Ontario, Canada 2001-2005 B.Sc. Hons. The University of Western Ontario London, Ontario, Canada 2005-2009 M.D. Queen's University Kingston, Ontario, Canada 2009-2011 Family Medicine Residency 2011-2012 Emergency Medicine Fellowship
Honours and Awards:	Schulich School of Medicine & Dentistry Bioethics Award 2009
Related Work Experience:	Physician: Family and Emergency Medicine Multiple locations 2011-2013

Publications:

Morgenstern J. Unspoken suffering: Obstacles to palliative care in pediatric medicine. *UWOMJ* 2006;74(2):4-6.

Morgenstern J. The medical oath: Honorable tradition or ancient ritual? *UWOMJ* 2008;78(1):27-9.